**Supporting Information for**

# **The Evolution of Research in Resources, Conservation & Recycling Revealed by Word2vec-enhanced Data Mining**

Jun-Jie Zhu1,2, Zhiyong Jason Ren1,2\*

1. Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA.
2. Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ 08544, USA.

\*Corresponding author; email: zjren@princeton.edu

**Summary**

Number of pages 38

Number of figures 9

Number of tables 12

Number of texts 6

Table S1. Acronyms that were identified (frequency ≥ 3) and their full descriptions (punctuations were removed and remain as singular form).

|  |  |  |  |
| --- | --- | --- | --- |
| **Full description** | **Acronym** | **Full description** | **Acronym** |
| Acrylonitrile Butadiene Styrene | ABS | Direct Reduced Iron | DRI |
| Advanced Dry Recovery | ADR | Differential Scanning Calorimetry | DSC |
| Analytic Hierarchy Process | AHP | Electric Arc Furnace | EAF |
| Acid Mine Drainage | AMD | Electrical and Electronic Equipment | EEE |
| Automobile Shredder Residue | ASR | Environmental Impact Assessment | EIA |
| Best Available Techniques | BAT | Eco Industrial Park | EIP |
| Business As Usual | BAU | Environmental Kuznets Curve | EKC |
| Building Information Modeling | BIM | Exergetic Life Cycle Assessment | ELCA |
| Biodegradable Municipal Waste | BMW | End of Life Vehicle | ELV |
| Biochemical Oxygen Demand | BOD | Environmentally weighted Material Consumption | EMC |
| Basic Oxygen Furnace | BOF | Environmental Product Declaration | EPD |
| Best Worst Method | BWM | Extended Producer Responsibility | EPR |
| Cost Benefit Analysis | CBA | Energy Return On Investment | EROI |
| California Bearing Ratio | CBR | Effluent Treatment Plant | ETP |
| Carbon Capture and Storage | CCS | Extended Theory of Planned Behavior | ETPB |
| Carbon Capture and Utilization | CCU | Economy Wide Material Flow Analysis | EWMFA |
| Carbon Capture, Utilization and Storage | CCUS | Granulated Blast Furnace Slag | GBFS |
| Clean Development Mechanism | CDM | Geologic Carbon Sequestration | GCS |
| Cation Exchange Capacity | CEC | Gross Domestic Product | GDP |
| Cumulative Energy Demand | CED | Glass Fibre Reinforced Plastic | GFRP |
| Cumulative Exergy Extracted from The Natural Environment | CEENE | Geographic Information System | GIS |
| Carbon Fibre Reinforced Polymers | CFRP | Green Supply Chain Management | GSCM |
| Computable General Equilibrium | CGE | Ground Tire Rubber | GTR |
| Closed Loop Supply Chain | CLSC | Grey Water Footprint | GWF |
| Controlled Low Strength Materials | CLSM | Global Warming Potential | GWP |
| Calcium Magnesium Acetate | CMA | High Density Poly Ethylene | HDPE |
| Community Multiscale Air Quality | CMAQ | Hot Mix Asphalt | HMA |
| Chemical Oxygen Demand | COD | Hydraulic Retention Time | HRT |
| Critical Raw Material | CRM | Household Solid Waste | HSW |
| Conservation Reserve Program | CRP | Internet Of Things | IOT |
| Cathode Ray Tube | CRT | Intergovernmental Panel on Climate Change | IPCC |
| Contingent Valuation Method | CVM | Interpretive Structural Modeling | ISM |
| Construction Waste Management | CWM | Indium Tin Oxide | ITO |
| Dissolved Air Flotation | DAF | Joint Forest Management | JFM |
| Data Envelopment Analysis | DEA | Knowledge, Attitudes, and Practice | KAP |
| Density Functional Theory | DFT | Key Performance Indicator | KPI |
| Decision Making Trial and Evaluation Laboratory | DMATEL | Life Cycle Assessment | LCA |
| Domestic Material Consumption | DMC | Life Cycle Costing | LCC |
| Direct Material Input | DMI | Liquid Crystal Display | LCD |
| Domestic Processed Output | DPO | Life Cycle Inventory | LCI |

|  |  |  |  |
| --- | --- | --- | --- |
| **Full description** | **Acronym** | **Full description** | **Acronym** |
| Life Cycle Impact Assessment | LCIA | Pearl River Delta | PRD |
| Light Emitting Diode | LED | Polyvinylchloride | PVC |
| Leadership In Energy and Environmental Design | LEED | Recycled Aggregate Concrete | RAC |
| Lime Kiln Dust | LKD | Reclaimed Asphalt Pavement | RAP |
| Logarithmic Mean Divisia Index | LMDI | Recycled Concrete Aggregate | RCA |
| Loss On Ignition | LOI | Recycled Concrete Powder | RCP |
| Mean Absolute Error | MAE | Refuse Derived Fuel | RDF |
| Mechanical Biological Treatment | MBT | Rare Earth Element | REE |
| Multi Criteria Decision Analysis | MCDA | Self-Compacting Concrete | SCC |
| Multi Criteria Decision Making | MCDM | Supply Chain Management | SCM |
| Material Circularity Indicator | MCI | Structural Decomposition Analysis | SDA |
| Material and Energy Flow Analysis | MEFA | Sustainable Development Goal | SDG |
| Material Flow Analysis | MFA | Scanning Electron Microscopy | SEM |
| Mixed Integer Linear Programming | MILP | Substance Flow Analysis | SFA |
| Multilayer Perceptron | MLP | Sewage Sludge Ash | SSA |
| Material Recovery Facility | MRF | Sustainable Supply Chain Management | SSCM |
| Multi Regional Input Output | MRIO | Solid State Fermentation | SSF |
| Municipal Solid Waste | MSW | Shared Socioeconomic Pathway | SSP |
| Municipal Solid Waste Incineration | MSWI | Solid Waste Management | SWM |
| Municipal Solid Waste Management | MSWM | Strength, Weakness, Opportunity, and Threat | SWOT |
| Norm Activation Model | NAM | Triple Bottom Line | TBL |
| Nonenergy Use Emission Accounting Table | NEAT | Toxicity Characteristic Leaching Procedure | TCLP |
| National Health Service | NHS | Total Material Requirement | TMR |
| National Industrial Symbiosis Programme | NISP | Theory of Planned Behavior | TPB |
| Net Present Value | NPV | Upflow Anaerobic Sludge Blanket | UASB |
| Non Energy Use | NEU | Volatile Fatty Acid | VFA |
| Non Renewable Energy Use | NREU | Waste Cooking Oil | WCO |
| Ordinary Portland Cement | OPC | Waste Electrical and Electronic Equipment | WEEE |
| Organic Waste Research | ORWARE | Water Energy Food | WEF |
| Pay As You Throw | PAYT | Waste Foundry Sand | WFS |
| Principal Component Analysis | PCA | Warm Mix Asphalt | WMA |
| Printed Circuit Board | PCB | Waste Printed Circuit Board | WPCB |
| Pro Environmental Behavior | PEB | Waste To Energy | WTE |
| Primary Energy Demand | PED | Wastewater Treatment Plant | WWTP |
| Platinum Group Metal | PGM | X Ray Diffraction | XRD |
| Polyhydroxyalkanoate | PHA | X Ray Fluorescence | XRF |
| Polyhydroxybutyrate | PHB | Yangtze River Economic Belt | YREB |
| Polylactic Acid | PLA |  |  |
| Polymer Matrix Composite | PMC |  |  |
| Palm Oil Mill Effluent | POME |  |  |

Table S2. Chemicals and metals that were identified (combined frequency ≥ 3) and unified with their varying forms (lowercase).

|  |  |
| --- | --- |
| Form 1 | Form 2 |
| carbon dioxide | co2, co(2) |
| carbon monoxide | co |
| nitrous oxide | n2o, n(2)o |
| sulfur dioxide | so2, so(2) |
| cd2+ | cadmium |
| li | lithium |
| as(v) | arsenate |

Table S3. Keywords (frequency ≥ 3) identified and their final replaced term (bold). Keywords may be listed as their singular forms in American English while the actual text replacement also included their plural forms or British English.

|  |  |
| --- | --- |
| No. | Keywords |
| 1 | agricultural residue, **agricultural waste** |
| 2 | anaerobic co digestion, co digestion, **codigestion** |
| 3 | analytical network process, **analytic network process** |
| 4 | beijing city, **beijing** |
| 5 | bibliometric, **bibliometric analysis** |
| 6 | big data analytic, **big data** |
| 7 | biogas production, **biogas** |
| 8 | bottom up approach, **bottom up modeling** |
| 9 | building material stock, **building material** |
| 10 | c&d waste, construction waste, demolition waste, construction and demolition waste recycling, construction and demolition waste management, **construction and demolition waste** |
| 11 | carbon capture, co2 capture, carbon storage, co2 storage, carbon utilization, co2 utilization, ccs, carbon capture and utilization or storage, carbon capture, utilization and storage, co2 capture and utilization or storage, co2 capture, utilisation and storage, **ccus** |
| 12 | cement paste, **cement mortar** |
| 13 | cement production, **cement manufacturing** |
| 14 | circular bioeconomy, **circular economy** |
| 15 | construction sector, **construction industry** |
| 16 | consumer attitude, **consumer behavior** |
| 17 | contingent valuation method, **contingent valuation** |
| 18 | copper scrap, **copper slag** |
| 19 | covid 19 pandemic, **covid 19** |
| 20 | decision making model, **decision making** |
| 21 | decision support system, **decision support** |
| 22 | deposit refund system, **deposit refund** |
| 23 | domestic rainwater harvesting, **rainwater harvesting** |
| 24 | domestic sewage, sewage, municipal wastewater, **wastewater** |
| 25 | driving factor, **driving force** |
| 26 | ecological footprint, environmental footprint, product environmental footprint, **footprint** |
| 27 | ecological sustainability, environmental sustainability, **sustainability** |
| 28 | economy wide material flow accounting, **economy wide material flow analysis** |
| 29 | electrical and electronic equipment, waste of electric and electronic equipment, reuse of electrical and electronic equipment, waste of electrical and electronic equipment, waste electric and electronic equipment, waste electronic and electrical equipment, waste from electrical and electronic equipment, weee management, **weee** |
| 30 | end of life allocation, end of life management, **end of life** |
| 31 | energy consumption, energy use, primary energy demand, **energy demand** |
| 32 | enhanced landfill mining, **landfill mining** |
| 33 | environmental impact analysis, impact assessment, impact analysis, **eia** |
| 34 | environmental input output analysis, input output model, input output table, **input output analysis** |
| 35 | environmental life cycle assessment, life cycle, **lca** |
| 36 | environmental pollution, **pollution** |
| 37 | environmental problem, **environmental issue** |
| 38 | final consumption, final demand, **consumption** |
| 39 | food energy water, water energy food, water energy food nexus, **food energy water nexus** |
| 40 | fuzzy ahp, analytical hierarchical process, analytical hierarchy process, **ahp** |
| 41 | granulated blast furnace slag, **blast furnace slag** |
| 42 | graywater recycling, **graywater reuse** |
| 43 | green supply chain practice, green supply chain, **gscm** |
| 44 | household food waste prevention, food waste prevention, household food waste, consumer food waste, **food waste** |
| 45 | incinerator bottom ash, **mswi bottom ash** |
| 46 | led lamp, **led** |
| 47 | li ion battery, lithium battery, **lithium ion battery** |
| 48 | lifespan analysis, **lifespan** |
| 49 | lithium recovery, **lithium recycling** |
| 50 | lmdi method, **lmdi** |
| 51 | logistic regression, **logistic model** |
| 52 | mainland china, eastern china, **china** |
| 53 | material and energy flow analysis, material and energy flow, **mefa** |
| 54 | material recycling, materials recovery, **material recovery** |
| 55 | material use, material utilization, **material consumption** |
| 56 | metal recovery, metals recycling, **metal recycling** |
| 57 | multi objective modeling, **multi objective optimization** |
| 58 | multiple criteria analysis, **multi criteria analysis** |
| 59 | nutrient recycling, **nutrient recovery** |
| 60 | old scrap, **scrap** |
| 61 | organic compound, **organic matter** |
| 62 | participation rate, **participation** |
| 63 | pet bottle, pet recycling, pet, **polyethylene terephthalate** |
| 64 | phosphorus recycling, **phosphorus recovery** |
| 65 | plastic recovery, **plastic recycling** |
| 66 | public acceptance, public awareness, public opinion, **public perception** |
| 67 | pulp and paper, **pulp and paper industry** |
| 68 | questionnaire survey, **questionnaire** |
| 69 | rainwater collection, **rainwater harvesting** |
| 70 | rare earth metal, rare earth, **rare earth element** |
| 71 | reclaimed water, **recycled water** |
| 72 | recycling efficiency rate, **recycling efficiency** |
| 73 | resource recycling, **resource recovery** |
| 74 | resource use, resource utilization, **resource consumption** |
| 75 | sewage sludge, municipal sewage sludge, wastewater sludge, **sludge** |
| 76 | social economic, socio economic, **socioeconomic** |
| 77 | social economy, socio economy, **socioeconomy** |
| 78 | societal metabolism, **social metabolism** |
| 79 | steel industry, **iron and steel industry** |
| 80 | steel scrap, steelmaking slag, steel scrap, **steel slag** |
| 81 | steelmaking, steel plant, **steel production** |
| 82 | strength weakness opportunity threat, **swot** |
| 83 | sustainable resources management, **sustainable resource management** |
| 84 | sustainable supply chain, **sscm** |
| 85 | system dynamic model, system dynamic, **system dynamics model** |
| 86 | techno economic analysis, **techno economic assessment** |
| 87 | urban agriculture, **agriculture** |
| 88 | urban industrial symbiosis, **industrial urban symbiosis** |
| 89 | urban residential building, **residential building** |
| 90 | waste recycling, waste reuse, **waste recovery** |
| 91 | water consumption, water utilization, **water use** |
| 92 | water recycling, **water reuse** |

Table S4. Other miscellaneous keywords (frequency ≥ 3) were combined with their final replaced term (bold). “XYZ” stands for any word(s).

|  |  |
| --- | --- |
| No. | Keywords |
| 1 | phosphate XYZ, **phosphorus XYZ** |
| 2 | plastics XYZ, **plastic XYZ** |
| 3 | nutrients XYZ, **nutrient XYZ** |
| 4 | metals XYZ, **metal XYZ** |
| 5 | behavioral XYZ, **behavior XYZ** |
| 6 | pollutant XYZ , **pollution XYZ** |
| 7 | resources XYZ, **resource XYZ** |
| 8 | licoo2 XYZ, **lithium XYZ** |
| 9 | benefit cost analysis, cost benefit, **cba** |
| 10 | black water, **blackwater** |
| 11 | characteristic, **characterization** |
| 12 | closedloop, **closed loop** |
| 13 | co operative, **cooperative** |
| 14 | doorstepping, **door stepping** |
| 15 | ecodesign, **eco design** |
| 16 | ecological efficiency, **eco efficiency** |
| 17 | flyash, **fly ash** |
| 18 | foodservice, **food service** |
| 19 | geovisualization, **geo visualization** |
| 20 | gray water, **graywater** |
| 21 | health care, **healthcare** |
| 22 | heavy metal ion, **heavy metal** |
| 23 | lifecycle, **life cycle** |
| 24 | lightweighting, **light weighting** |
| 25 | make up, **makeup** |
| 26 | polyvinylchloride, **polyvinyl chloride** |
| 27 | pre processing, **preprocessing** |
| 28 | pre treatment, **pretreatment** |
| 29 | waste water, **wastewater** |
| 30 | wastepaper, **waste paper** |
| 31 | wastepicker, **waste picker** |

**Text S1. Title-based keywords generation**

In addition to author keywords, title was also used to generate additional meaningful keywords by tokenizing to *n*-gram (or terms consisted of *n* adjacent words; *n* = 1, 2, 3, 4, and 5). Only particular generated-keywords were retained if they were found in the library of preprocessed author keywords (without stemming; frequency > 1). To minimize the noises, single generated-keywords were first moved if they were stop-words (common words, such as *that*, *the*, *there*, and *is*). Furthermore, to avoid a repeating count, only the generated-keywords with a unique meaning in the same paper were included (generated-keywords that had full/partial repeated words with author keywords in the same paper were excluded). Equally important, British English was converted to American English at the very beginning.

**Text S2. The six-step approach for Word2vec modeling**

Six fields of data (*year*, *date*, *author keywords*, *title*, *abstract*, and *citation*) were used to conduct the Word2vec-based investigations using a Python package *gensim* (Gensim, 2022). In the step of “word tokenization and phrases”, title and abstract were tokenized by *unigram* (single word) based on batches of sentences; punctuations (such as hyphen, parentheses, brackets, braces, apostrophe, and quotation marks) were replaced with empty space to reduce the noises. A post-hoc correction was made to correct three high frequent keywords (*industry 3.5*, *industry 4.0*, *pm2.5*) due to incorrect removal of period. Word phrases was then applied to connect *n* adjacent words (*n* = 1, 2, 3, 4) and retain frequent terminologies based on two parameters, *min\_count* (the minimum frequency) and *threshold* (a score that is used along with *min\_count* to screen phrases; detailed in *gensim* document, Gensim (2022)); the two parameters were determined in a later step. The above pretreatment was only applied to title and abstract, while author keywords remained their original forms before combining them together for the “terminology preprocessing” in the second step. Specifically, the step was mainly performed based on the six-step deep word preprocessing approach as introduced previously, to minimize the textual noises and combine acronyms or synonyms. In addition to the original deep preprocessing, additional acronyms were identified from abstracts and corrected in the step 3, and new acronyms were also fed back to the keyword preprocessing in the section 2.1 to enhance the treatment. Furthermore, processing of abstracts generated many meaningless text corpuses, so additional stop-words treatment and corrections were applied in the step 4 using an iterative, trail-and-error inspection approach. In general, the first four steps helped to prepare a relatively clean and meaningful corpus of textual data, the next two steps were used to build the model and improve the model reliability.

The two parameters *min\_count* and *threshold* were used to control permitted terms; the higher their values are, less terminologies are permitted to be included in the model but with less noises (meaningless terms). In contrast, the lower their values are, the more terminologies are included in the analysis but with more noises. We therefore conducted a retrievability assessment to understand their effects on the retrievability (or recall) of keywords. We tested their values from 1 to 10 while other model parameters remained as constants and used processed keywords that have a frequency more than *f* (= 3, 4, or 5) to assess the percentage of keywords that could be retrieved (Fig. S1). The results show that the pattern of retrievability matches well with the tested frequency. For example, the retrievability for keyword frequency of 4 decreases rapidly when *min\_count* is above 4. We finally adopted [3, 2] as to be the values of [*min\_count*, *threshold*], which helped to retrieve 99.5%, 99.9%, and 100% of keywords that have a frequency of at least 3, 4, and 5, respectively.

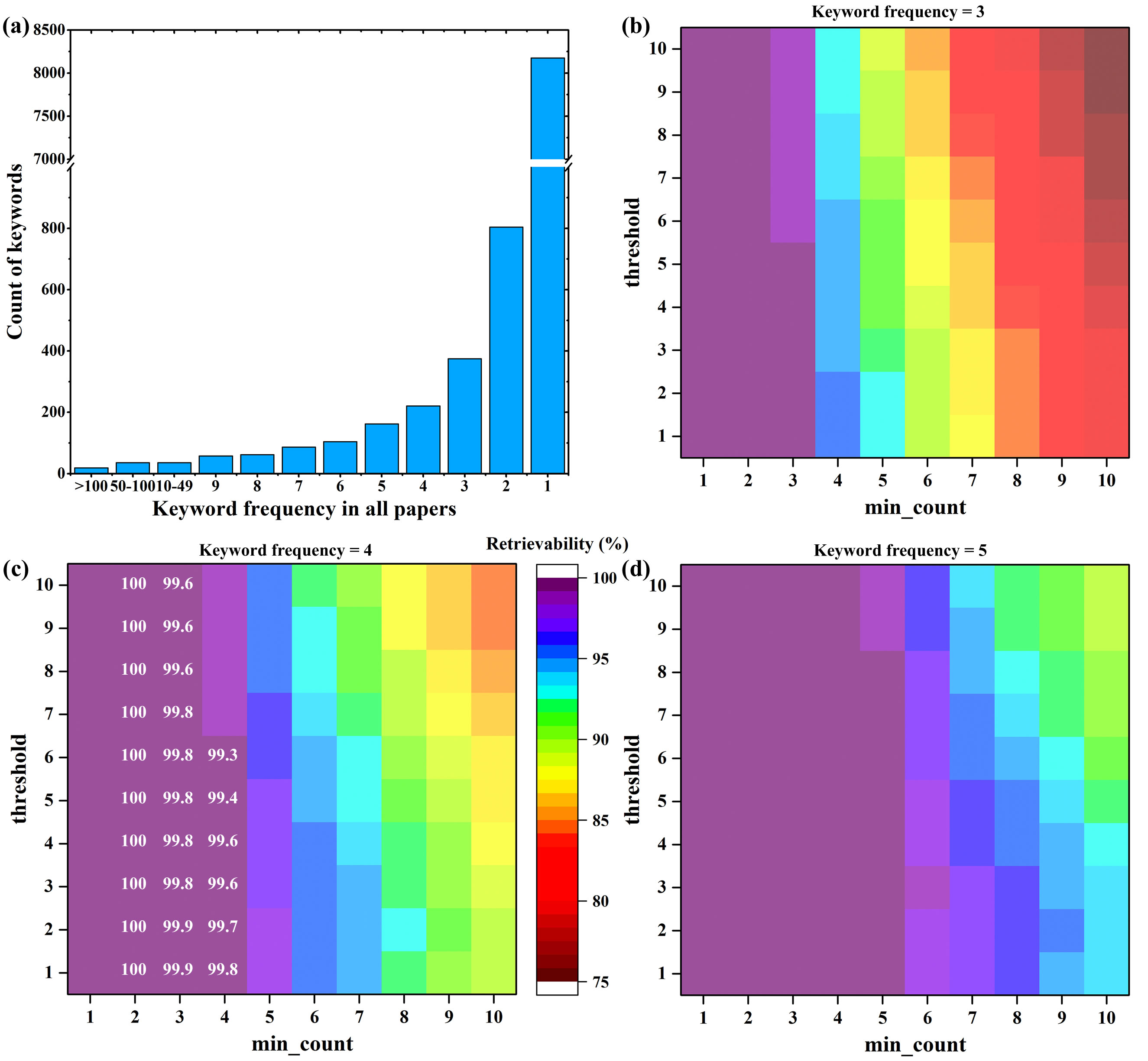


Fig. S1. Retrievability assessment in the word2vec modeling. (a) Count of keywords based on their respective frequencies in all RCR papers. Heat map plots of the word2vec model retrievability for keyword frequency that is equal to or more than (b) 3, (c) 4, and (d) 5. Same color can have minor different levels of retrievability, and some details are labeled in the plot of frequency of 4.

The last step was to optimize the Word2vec model based on several important model hyperparameters, including *vector\_size* (number of vectors used), *window* (the maximum distance to be analyzed from the current word to a target word), *epochs* (number of iterations for learning over the textual data), and *alpha* (the initial learning rate). A problem was that the default Word2vec model was not designed for supervised learning, so we prepared two lists of terminologies and added a similarity evaluation step to monitor the model performance and guide the model optimization. The first list (Table S5) included 40 pairs of synonyms (e.g., *climate change* and *global warming*) or terms that are tightly connected (e.g., *struvite* and *phosphorus*) in the scope of RCR or environmental engineering. The second list (Table S6) included 40 pairs of terms that are scientifically irrelevant (e.g., *zinc* and *water conservation*), especially in the scope of RCR. To quantify the model performance, each of the paired terms was assessed their similarity based on different combinations of model hyperparameters, and the overall similarity (*sim(α, β)overall*) was calculated by averaging of the similarities for the respective 40 pairs of terms (eq.(5); the equation index is kept consistently with the main text). The objective function was the difference between the overall similarities of the two lists, or called as similarity gap (eq.(6)). A similarity value between two terms can range from -1 (opposite terms) to 1 (identical terms), but it typically ranges from 0 to 1 because all the terms are from papers that focused on a relatively small research area (RCR); and a better model should have a bigger similarity gap, which has a similar range.

(5)

(6)

We tested 600 combinations of the four hyperparameters in the optimization step and determined the optimum model with *vector\_size* of 250, *window* of 12, *epochs* of 10, and *alpha* of 0.005 (Fig. S2). The optimum model was able to achieve the maximum similarity gap of 0.866 between the lists 1 (overall similarity ≈ 0.850) and 2 (≈ -0.015). As a result, the model was ready to be the final Word2vec-RCR model for the subsequent analyses. To visualize the model, we adopted a t-distributed stochastic neighbor embedding (t-SNE) method (Hinton and Roweis, 2002; van der Maaten and Hinton, 2008) to compress the original 250 dimensions to a two-dimensional plot based on an initialization embedding of principal component analysis (PCA) (perplexity=50) using Python package *scikit-learn* and module *manifold* (Sklearn, 2022).

Table S5. The list of 40 pairs of keywords (stemmed form) that are conceptually close.

|  |  |
| --- | --- |
| **Keyword 1** | **Keyword 2** |
| *agricultur* | *crop* |
| *ahp* | *mcda* |
| *air qual* | *air pollut* |
| *anaerobic digest* | *bioga* |
| *biodiesel* | *bioenergi* |
| *bottom ash* | *combust* |
| *carbon emiss* | *carbon footprint* |
| *cba* | *economic analysi* |
| *circular economi* | *circular business model* |
| *cleaner product* | *pollution prevent* |
| *climate chang* | *global warm* |
| *clsc* | *gscm* |
| *consumer behavior* | *willingness to pay* |
| *copper* | *zinc* |
| *economic analysi* | *economic assess* |
| *electric vehicl* | *lithium ion batteri* |
| *energi* | *exergi* |
| *energi* | *energy demand* |
| *energy conserv* | *energy sav* |
| *food wast* | *food loss* |
| *ghg* | *emission reduct* |
| *gwp* | *ghg* |
| *lca* | *lcc* |
| *lcc* | *lci* |
| *mcda* | *mcdm* |
| *metal* | *copper* |
| *metal recycl* | *hydrometallurgi* |
| *mfa* | *sfa* |
| *nutrient recoveri* | *phosphorus recoveri* |
| *pavement* | *road construct* |
| *power gener* | *electricity gener* |
| *pyrolysi* | *thermal treat* |
| *solid wast* | *msw* |
| *struvit* | *phosphorus* |
| *urban metabol* | *urban sustain* |
| *urban min* | *mine* |
| *waste manag* | *municipal wast* |
| *wastewat* | *wwtp* |
| *water conserv* | *water sav* |
| *water scarc* | *water reus* |

Table S6. The list of 40 pairs of keywords (stemmed form) that are conceptually distinct.

|  |  |
| --- | --- |
| **Keyword 1** | **Keyword 2** |
| *ahp* | *cadmium* |
| *aluminum* | *irrig* |
| *anaerobic digest* | *cobalt* |
| *behavior chang* | *steel product* |
| *biomass* | *social norm* |
| *bottom ash* | *water resourc* |
| *build* | *heavy met* |
| *circular economi* | *densiti* |
| *climate chang* | *scrap* |
| *compost* | *supply chain* |
| *compressive strength* | *trade* |
| *crop* | *disassembl* |
| *decision mak* | *dust* |
| *economic growth* | *pretreat* |
| *electric vehicl* | *soil* |
| *electronic wast* | *irrig* |
| *exergi* | *recycling behavior* |
| *food wast* | *copper* |
| *ghg* | *mechan* |
| *gscm* | *ammonia* |
| *international trad* | *digest* |
| *material stock* | *anaerob* |
| *pcb* | *air pollut* |
| *plastic* | *rainwater harvest* |
| *plastic wast* | *water sav* |
| *polyethylene terephthal* | *water suppli* |
| *polym* | *water scarc* |
| *pyrolysi* | *economic develop* |
| *remanufactur* | *organic matt* |
| *reverse logist* | *manur* |
| *sharing economi* | *ammonia* |
| *sludg* | *tpb* |
| *supply chain* | *ash* |
| *swm* | *densiti* |
| *system dynamics model* | *anaerob* |
| *water footprint* | *slag* |
| *water us* | *wpcb* |
| *weee* | *crop* |
| *wpcb* | *land us* |
| *zinc* | *water conserv* |

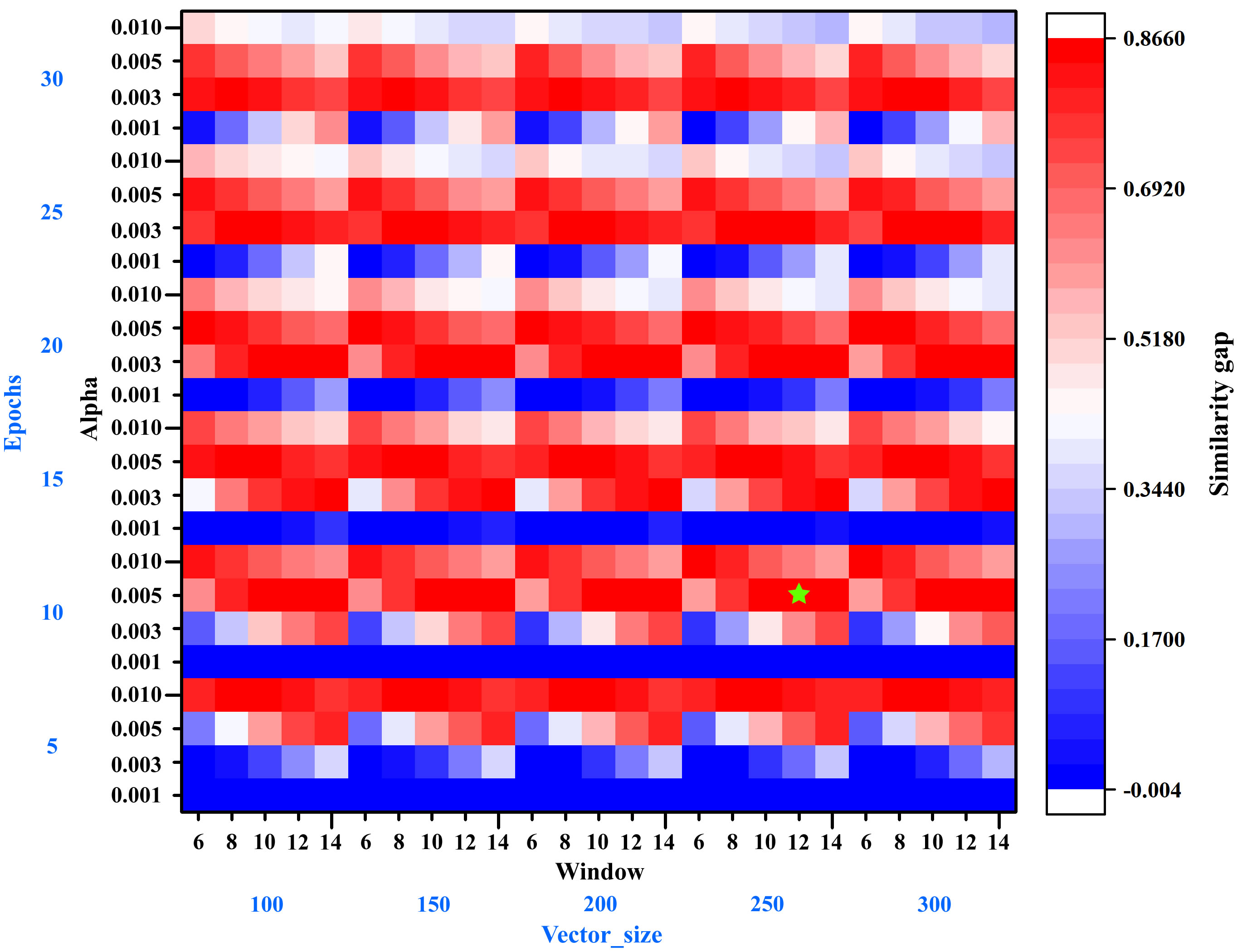


Fig. S2 Heat map plot of similarity gap based on different combinations of hyperparameters, including 5 *vector\_size* (100, 150, 200, 250, and 300), 5 *window* (6, 8, 10, 12, and 14), 6 *epochs* (5, 10, 15, 20, 25, and 30), and 4 *alpha* (0.001, 0.003, 0.005, and 0.01). The green star mark shows where the optimized model is located.

**Text S3. Word2vec-based resource-surrogates labeling method**

We first initialized three *resource keywords* (e.g., *air*, *air quality*, and *air pollution*) that were used to represent each of the 10 resources (Table S7). Second, we screened the top similar terminologies, from the Word2vec-RCR model (frequency ≥ 10), for each resource by ranking their resource-related similarities. We only focused on terminologies that are in the library of author keywords with a minimum frequency of 3 to avoid meaningless words. It is worth to note that the preprocessing removed non-technical “paper” terms, such as “this paper”, “the paper”, “present paper”, and “that paper”, to minimize the confusion of “paper” as the physical material for the model. The top 60 similar terminologies were used in a resource-based topical association analysis, which helped to understand interconnections among the resource-based research. To reveal the focus of all publications with respect to resource research, we then identified resource-surrogates from the terminology-candidates and used these surrogates to label the publications (Fig. S3). Specifically, we screened the top candidates with a similarity that was no less than 0.765 (90% × 0.85 = 0.765), and secured *p* resource-surrogates from the candidates based on expert knowledge. The next step was to calculate a *publication similarity* score (*sim(A, B)publication*) with the respective resource-surrogates (*Ai*; *i* = 1, 2, …, and *p*) and preprocessed terms (*Bj*) from its corresponding textual corpus for each resource and each paper (eq.(7)). Giving ranked preprocessed terms (*Bj*; *j* = 1, 2, …, *q, …*, and *Q*) based on the similarity with resource-surrogates, it is possible to use any *q* value that ranges from 1 (only the most similar term is used) to *Q* (all preprocessed terms are used).

(7)

To determine the most appropriate *q* value, we manually reviewed 112 highly cited papers (title, keywords, and abstract) and labeled them with their studied resource(s) based on expert knowledge. The 112 papers were composed of 50 highest cited papers for all the time, another 50 highest cited papers but since 2015, and 12 additional highly cited papers (used to ensure a minimum five papers for each resource). On the other hand, for different *q* values (*q* = 2, 4, 6, …, and 200), corresponding publication similarity scores were calculated; meanwhile, a range of threshold values (0.7, 0.75, 0.8, 0.85, and 0.9) were tested to identify the highest classification accuracy based on different *q* and threshold values. A conventional accuracy could be biased because the classification matrix is relatively sparse with many zero values. Alternatively, a weighted average paper-wise accuracy was calculated. Specifically, a weighted classification accuracy for each paper (*Accuracypublication*) was first calculated by averaging the accuracies of true (the resources that were studied by the paper) and false (the resources that were not studied by the paper) classifications (eq.(8)). The overall accuracy was then calculated by averaging the individual accuracies of all papers (*n* = 112) (eq.(9)). Finally, once the optimum *q* and threshold values were determined, we applied the method to classify all publications with respect to all 10 resources.

(8)

(9)

Table S7. The 10 resources and their respective three initial resource keywords

|  |  |
| --- | --- |
| **Resource** | **Initial resource keywords** |
| Air | *air* |
| *air pollution* |
| *air quality* |
| Ecosystem | *ecosystem* |
| *ecosystem service* |
| *biodiversity* |
| Energy | *energy* |
| *energy demand* |
| *energy recovery* |
| Food | *food* |
| *food waste* |
| *food loss* |
| Land | *land* |
| *land use* |
| *land footprint* |
| Metals | *metal* |
| *metal recycling* |
| *hydrometallurgy* |
| Minerals | *mineral* |
| *mineral resource* |
| *mineral depletion* |
| Waste | *waste* |
| *waste management* |
| *solid waste* |
| Water | *water* |
| *wastewater* |
| *rainwater harvesting* |
| Wood | *wood* |
| *wood production* |
| *recovered wood* |

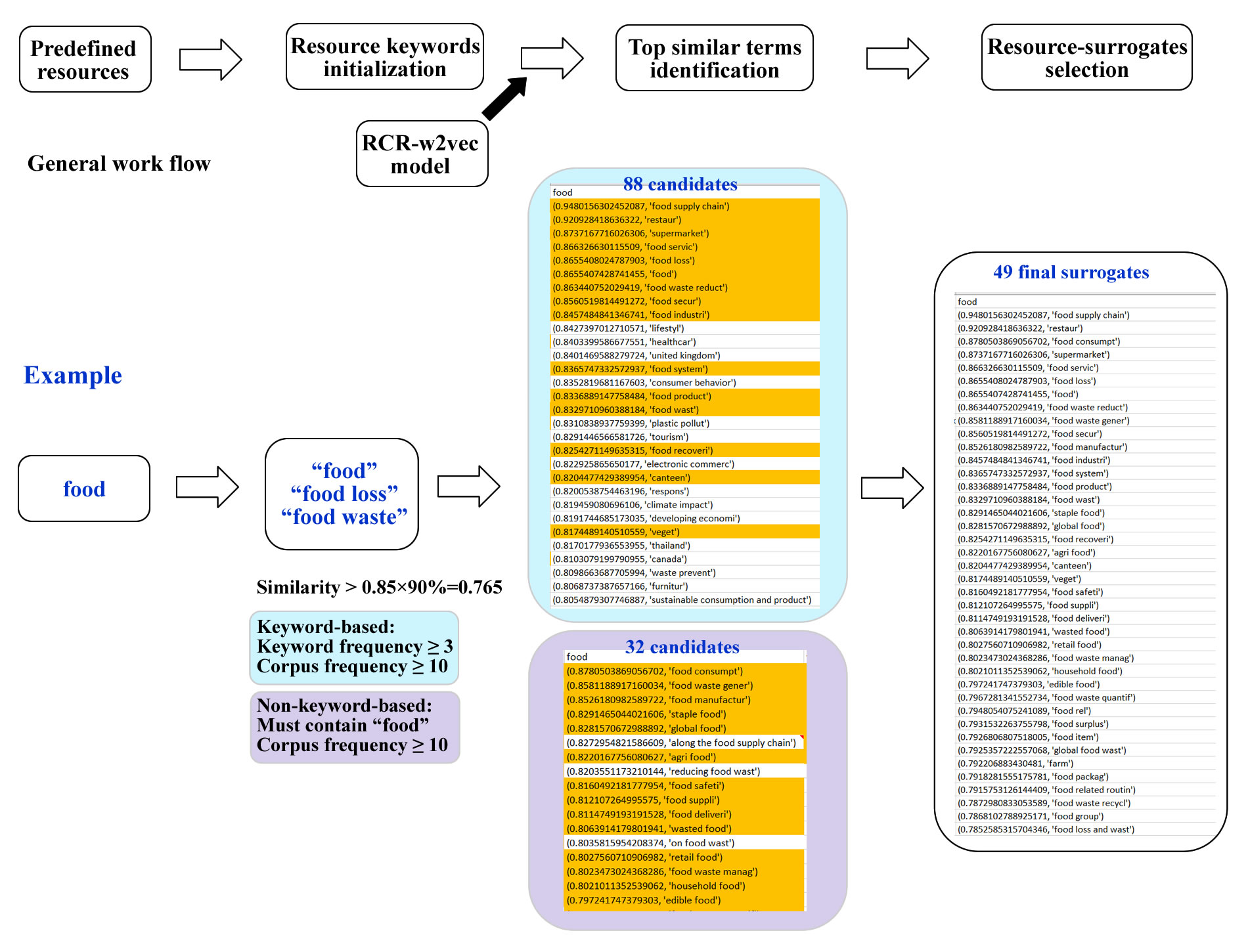


Fig. S. An example of resource surrogates identification for food based on the RCR-word2vec model. Two types (keyword-based and non-keyword-based) of terminology-candidates were identified. The keyword-based candidates were detected using the three defined resource keywords based on the RCR-word2vec model. The non-keyword-based candidates were revealed using word “food”, so any other high-similarity terminologies that contained “food” were inventoried. The final surrogates were determined based on expert knowledge while irrelevant, less representative, or inappropriate corpus were excluded from the candidates.

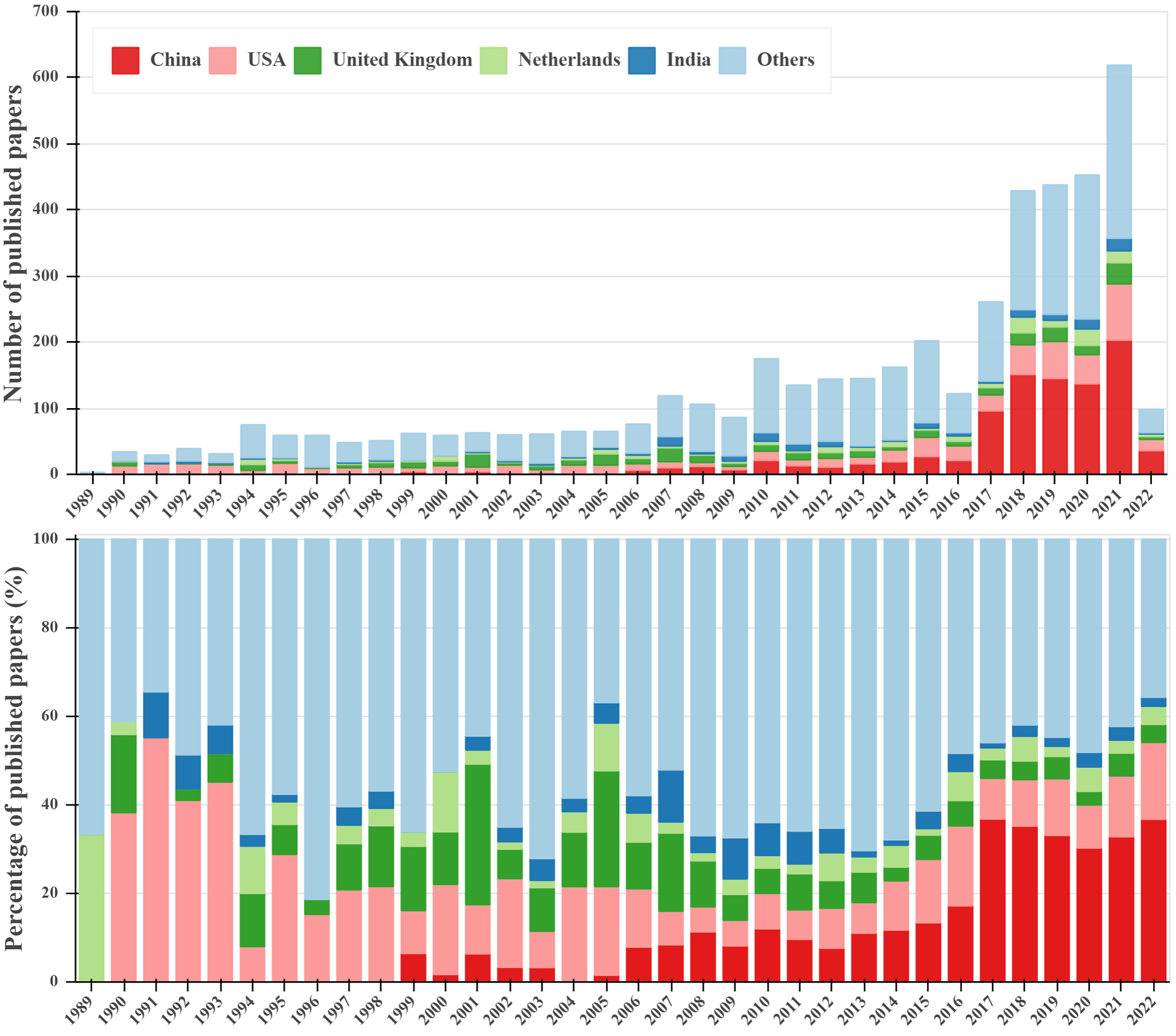


Fig. S4. Temporal and geospatial variations of research articles and reviews in RCR from 1989 to 2022 (web of science data updated to March, 2022). (a) Number of papers; (b) percentage of papers.

Table S8. Top 100 frequent keywords (lowercased, stemmed form) and their frequencies.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. | Keyword | Freq. | No. | Keyword | Freq. | No. | Keyword | Freq. |
| 1 | *recycl* | 793 | 35 | *supply chain* | 70 | 68 | *developing countri* | 43 |
| 2 | *lca* | 564 | 36 | *sustainable develop* | 67 | 68 | *carbon footprint* | 43 |
| 3 | *china* | 486 | 36 | *electronic wast* | 67 | 71 | *pollut* | 42 |
| 4 | *circular economi* | 311 | 38 | *optim* | 64 | 72 | *uncertainti* | 41 |
| 5 | *sustain* | 294 | 39 | *material flow* | 61 | 72 | *agricultur* | 41 |
| 6 | *wast* | 240 | 40 | *polici* | 60 | 74 | *consumpt* | 40 |
| 7 | *waste manag* | 233 | 41 | *build* | 59 | 74 | *bioga* | 40 |
| 8 | *mfa* | 219 | 42 | *emiss* | 58 | 76 | *consum* | 39 |
| 9 | *model* | 190 | 42 | *behavior* | 58 | 76 | *indic* | 39 |
| 10 | *msw* | 188 | 44 | *sludg* | 55 | 76 | *input output analysi* | 39 |
| 11 | *eia* | 163 | 45 | *resource recoveri* | 53 | 79 | *mswm* | 38 |
| 12 | *solid wast* | 146 | 46 | *india* | 52 | 79 | *barrier* | 38 |
| 13 | *municip* | 136 | 46 | *landfil* | 52 | 79 | *tpb* | 38 |
| 14 | *reus* | 135 | 48 | *anaerobic digest* | 51 | 82 | *decision mak* | 37 |
| 15 | *ghg* | 118 | 48 | *inciner* | 51 | 82 | *cement* | 37 |
| 16 | *food wast* | 116 | 50 | *fly ash* | 50 | 82 | *attitud* | 37 |
| 17 | *recoveri* | 115 | 50 | *construct* | 50 | 85 | *heavy met* | 36 |
| 18 | *swm* | 104 | 50 | *biomass* | 50 | 85 | *energy demand* | 36 |
| 19 | *resourc* | 93 | 50 | *lithium ion batteri* | 50 | 85 | *waste recoveri* | 36 |
| 20 | *wastewat* | 92 | 54 | *resource effici* | 49 | 85 | *aluminum* | 36 |
| 21 | *weee* | 90 | 54 | *sfa* | 49 | 85 | *brazil* | 36 |
| 22 | *carbon emiss* | 85 | 56 | *phosphorus* | 47 | 90 | *scenario* | 35 |
| 23 | *util* | 84 | 56 | *cost* | 47 | 90 | *rainwater harvest* | 35 |
| 23 | *energi* | 84 | 58 | *copper* | 46 | 90 | *ree* | 35 |
| 25 | *concret* | 81 | 58 | *system dynamics model* | 46 | 93 | *water* | 34 |
| 26 | *household* | 79 | 58 | *electric vehicl* | 46 | 93 | *footprint* | 34 |
| 27 | *plastic* | 78 | 61 | *industrial ecolog* | 45 | 95 | *material effici* | 33 |
| 28 | *united st* | 77 | 61 | *climate chang* | 45 | 95 | *methodolog* | 33 |
| 29 | *united kingdom* | 76 | 63 | *character* | 44 | 95 | *oil* | 33 |
| 30 | *compost* | 75 | 63 | *composit* | 44 | 95 | *aggreg* | 33 |
| 31 | *construction and demolition wast* | 74 | 63 | *carbon dioxid* | 44 | 95 | *scm* | 33 |
| 32 | *metal* | 73 | 63 | *wastewater treat* | 44 | 100 | *transport* | 32 |
| 32 | *end of lif* | 73 | 63 | *reverse logist* | 44 | 100 | *lcc* | 32 |
| 34 | *urban* | 70 | 68 | *effici* | 43 |  | | |

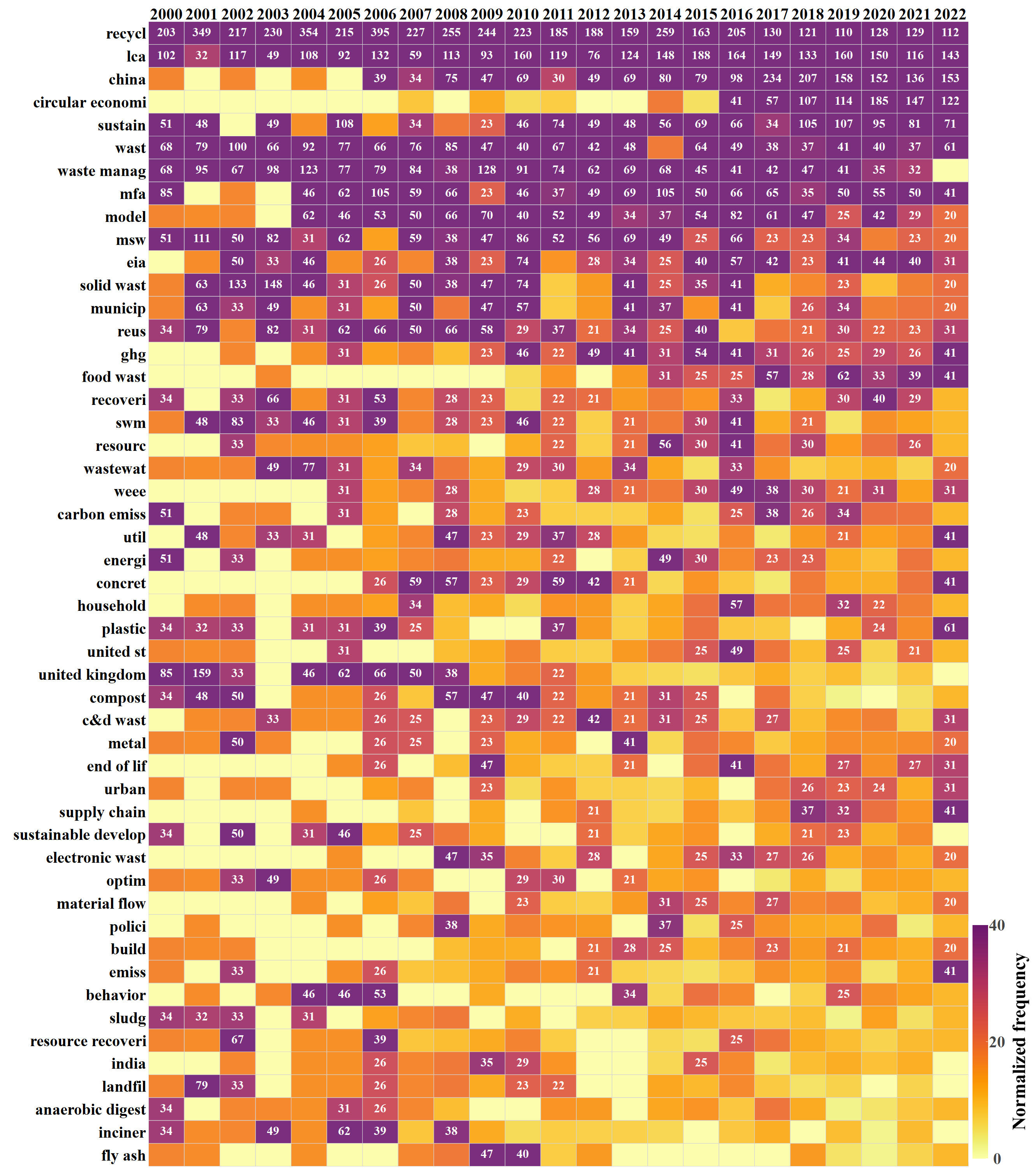


Fig. S5. Temporal trend of the top 50 frequent keywords based on normalized annual frequency. Higher frequencies (≥ 20) are labeled.

**Text S4. Annual top frequent keywords**

Annual top five frequent keywords over the 34 years are summarized in Table S9; *recycl* appeared the most common topic (22 years) that ranked at the first place, followed by *lca*, *china*, and *circular economi*. Other frequent keywords in all five top places include *wast* (frequency = 13), *waste manag* (12), *sustain* (8), and *msw* (8). Some keywords were studied together for common research interests or having an interaction to realize an objective; and most of the top co-occurring keywords can be expected, such as *msw* and *municip* (frequency = 103), *eia* and *lca* (94), and *lca* and *recycl* (85) (Table S10).

Table S9. Annual top five frequent keywords from 1989 to 2022 (March)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | Top frequent keywords | | | | |
| 1 | 2 | 3 | 4 | 5 |
| 1989 | *solid wast* | *composit* | *recycl* | *forecast* | *resource recoveri* |
| 1990 | *municip* | *solid wast* | *msw* | *coal* | *wast* |
| 1991 | *coal* | *oil* | *wast* | *model* | *recycl* |
| 1992 | *wast* | *calcium* | *magnesium* | *recoveri* | *urban* |
| 1993 | *recycl* | *solid wast* | *united st* | *msw* | *swm* |
| 1994 | *recycl* | *wast* | *metal* | *wastewat* | *energi* |
| 1995 | *recycl* | *lca* | *wast* | *plastic* | *fly ash* |
| 1996 | *pollut* | *wastewat* | *msw* | *municip* | *recycl* |
| 1997 | *recycl* | *lca* | *model* | *tanzania* | *waste manag* |
| 1998 | *recycl* | *msw* | *swm* | *optim* | *recoveri* |
| 1999 | *recycl* | *wast* | *wastewat* | *reus* | *model* |
| 2000 | *recycl* | *lca* | *mfa* | *united kingdom* | *wast* |
| 2001 | *recycl* | *united kingdom* | *msw* | *waste manag* | *reus* |
| 2002 | *recycl* | *solid wast* | *lca* | *wast* | *swm* |
| 2003 | *recycl* | *solid wast* | *waste manag* | *msw* | *reus* |
| 2004 | *recycl* | *waste manag* | *lca* | *wast* | *wastewat* |
| 2005 | *recycl* | *sustain* | *neu* | *lca* | *fossil fuel* |
| 2006 | *recycl* | *lca* | *mfa* | *waste manag* | *united kingdom* |
| 2007 | *recycl* | *waste manag* | *wast* | *lca* | *mfa* |
| 2008 | *recycl* | *lca* | *wast* | *china* | *model* |
| 2009 | *recycl* | *waste manag* | *lca* | *model* | *reus* |
| 2010 | *recycl* | *lca* | *waste manag* | *msw* | *solid wast* |
| 2011 | *recycl* | *lca* | *waste manag* | *sustain* | *wast* |
| 2012 | *recycl* | *lca* | *waste manag* | *msw* | *china* |
| 2013 | *recycl* | *lca* | *china* | *waste manag* | *mfa* |
| 2014 | *recycl* | *lca* | *mfa* | *china* | *waste manag* |
| 2015 | *lca* | *recycl* | *china* | *sustain* | *wast* |
| 2016 | *recycl* | *lca* | *china* | *model* | *mfa* |
| 2017 | *china* | *lca* | *recycl* | *mfa* | *model* |
| 2018 | *china* | *lca* | *recycl* | *circular economi* | *sustain* |
| 2019 | *lca* | *china* | *circular economi* | *recycl* | *sustain* |
| 2020 | *circular economi* | *china* | *lca* | *recycl* | *sustain* |
| 2021 | *circular economi* | *china* | *recycl* | *lca* | *sustain* |
| 2022 | *china* | *lca* | *circular economi* | *recycl* | *sustain* |

Table S10. Summary of the 31 couples of high frequent (≥ 30) co-occurring keywords.

|  |  |  |  |
| --- | --- | --- | --- |
| Order | Keyword 1 | Keyword 2 | Frequency |
| 1 | *msw* | *municip* | 103 |
| 2 | *eia* | *lca* | 94 |
| 3 | *lca* | *recycl* | 85 |
| 4 | *msw* | *solid wast* | 77 |
| 5 | *municip* | *solid wast* | 76 |
| 6 | *recycl* | *wast* | 71 |
| 7 | *recycl* | *reus* | 64 |
| 8 | *circular economi* | *recycl* | 61 |
| 9 | *recycl* | *waste manag* | 61 |
| 10 | *circular economi* | *lca* | 61 |
| 11 | *mfa* | *recycl* | 52 |
| 12 | *ghg* | *lca* | 51 |
| 13 | *lca* | *sustain* | 46 |
| 14 | *lca* | *waste manag* | 44 |
| 15 | *recycl* | *sustain* | 41 |
| 16 | *solid wast* | *swm* | 41 |
| 17 | *model* | *recycl* | 40 |
| 18 | *circular economi* | *sustain* | 39 |
| 19 | *material flow* | *mfa* | 39 |
| 20 | *china* | *lca* | 38 |
| 21 | *msw* | *swm* | 36 |
| 22 | *plastic* | *recycl* | 36 |
| 23 | *msw* | *recycl* | 35 |
| 24 | *china* | *recycl* | 35 |
| 25 | *recoveri* | *recycl* | 35 |
| 26 | *carbon emiss* | *china* | 32 |
| 27 | *msw* | *mswm* | 32 |
| 28 | *recycl* | *solid wast* | 31 |
| 29 | *mswm* | *swm* | 31 |
| 30 | *municip* | *swm* | 30 |
| 31 | *lca* | *mfa* | 30 |

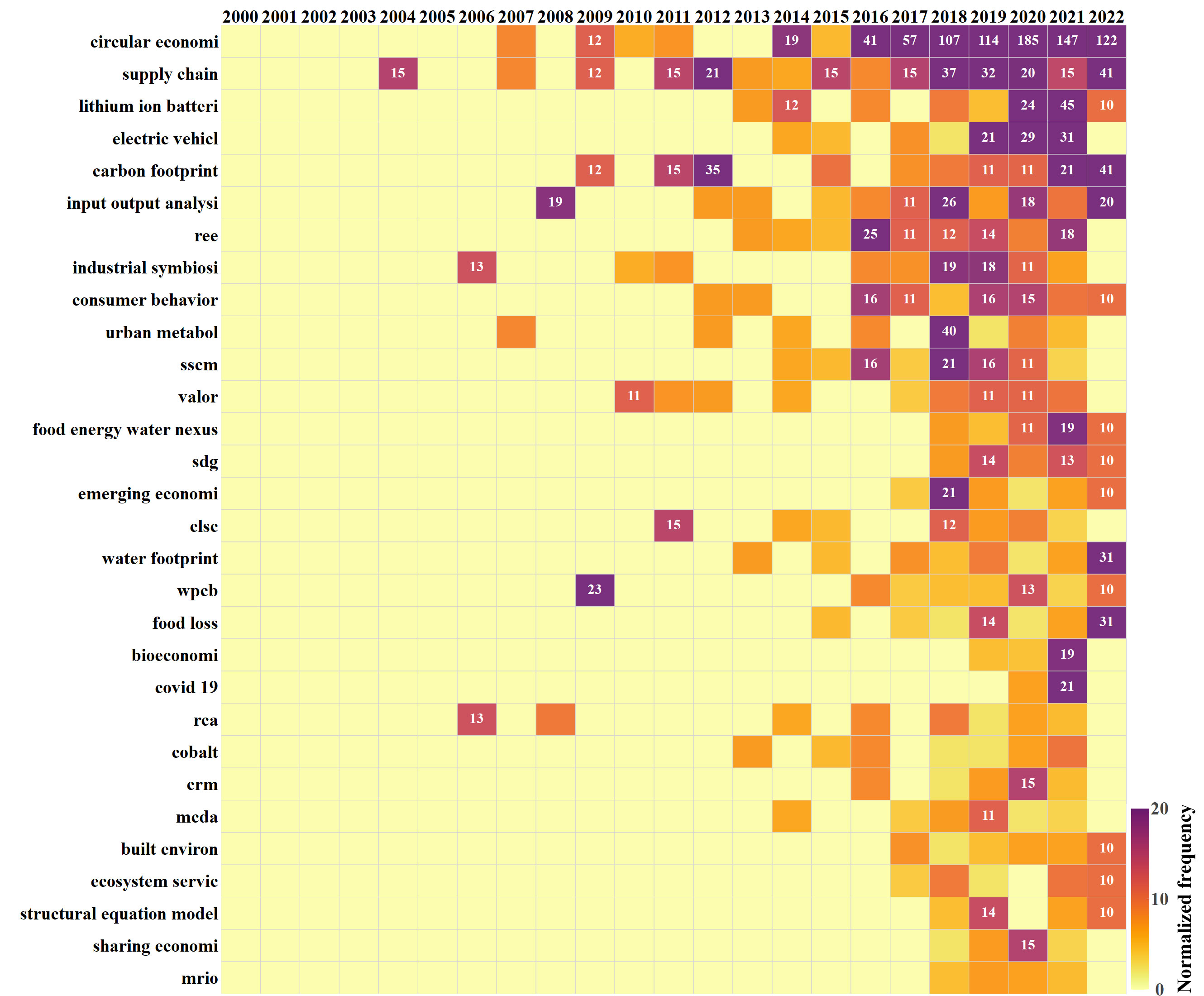


Fig. S6. Heat map plot for the temporal trend of the 30 emerging keywords based on normalized annual frequency; higher frequencies (≥ 10) are labeled.

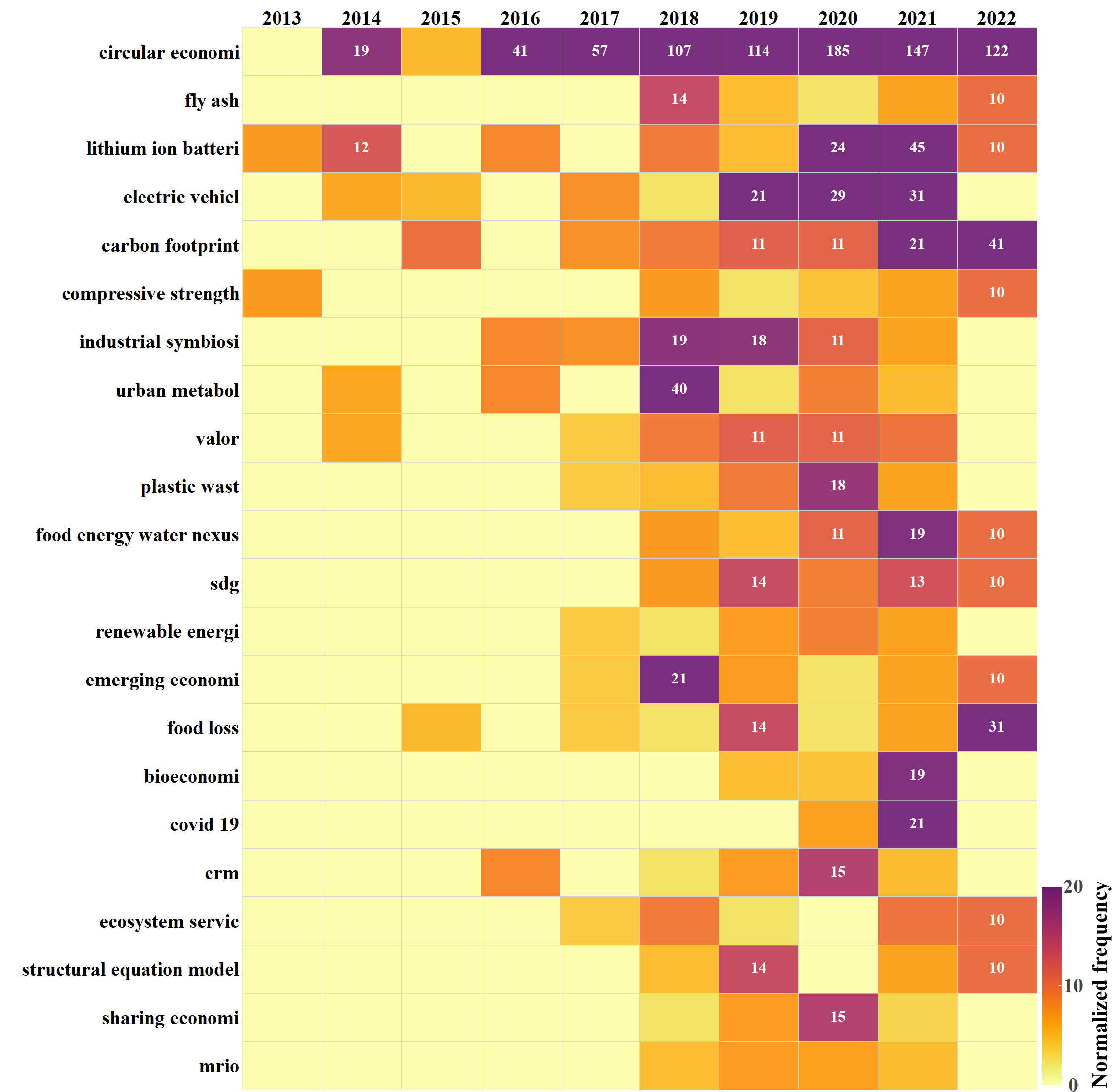


Fig. S7. Heat map plot for the temporal trend of the 22 emerging keywords using the most recent 10 years of data (2013-2017 vs. 2018-2022) based on normalized annual frequency; higher frequencies (≥ 10) are labeled.

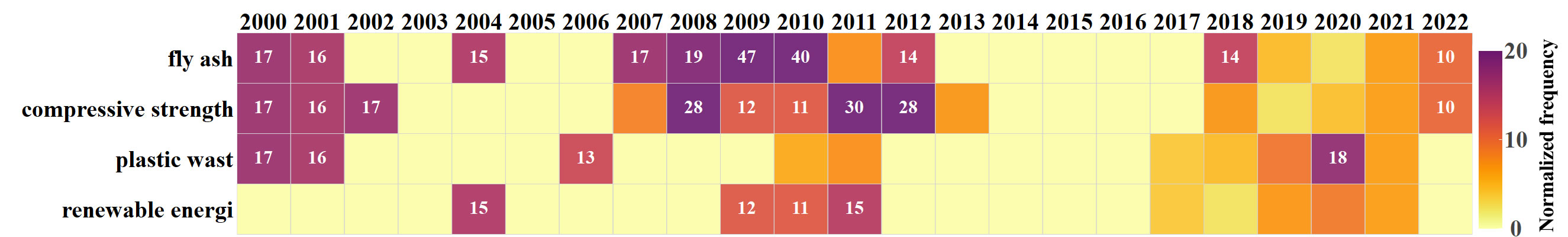


Fig. S. Heat map plot (2000-2022) for the temporal trend of the 4 “newly identified”, emerging keywords using the most recent 10 years of data (2013-2017 vs. 2018-2022) based on normalized annual frequency; higher frequencies (≥ 10) are labeled.

**Text S5. Similarity analysis of the associated top 60 similar terminologies for the 10 resources**

Among the 10 resources, the highest average similarity of the top terms are found in “minerals” (sim ≈ 0.973) and “wood” (0.951), indicating that their relevant research activities were more specific, so the associated topic-vectors are in regions with more dense neighbors. For example, similarity of the top terms to “minerals” ranges from 0.992 (*non fuel miner*) to 0.961 (*stakeholder analysi*), while similarity values for “wood” range from 0.971 (*energy cont*) to 0.935 (*beneficial us*), respectively (Table S11). In contrast, “food” (0.823), “water” (0.824), and “waste” (sim ≈ 0.824), have the least levels of average similarity. This is likely because they are more general concepts and have been discussed ubiquitously, which dilutes the density of associated topics near the resources in the vector space. Nonetheless, the top terms for the three resources are well in agreement with expert knowledge. For example, the top three terms for “waste” are *integrated waste manag* (0.884), *swm* (0.870), and *waste recoveri* (0.868). Even the least three topics, *solid waste recycl* (0.800), *waste collect* (0.800), and *landfill direct* (0.799), are closely associated with “waste”. Interestingly, several top terminologies reflect the research scopes of RCR according to the respective resources, such as “air” (*urban agglomer*), “ecosystem” (*urban sustain*), “energy” (*wte*), and “land” (*carbon sequestr*).

Table S11. Lists of the top 60 similar terminologies for each of the 10 resources.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Air | | Ecosystem | | Energy | |
| term | sim | term | sim | term | sim |
| *air* | – | *ecosystem* | – | *energi* | – |
| *air pollut* | – | *ecosystem servic* | – | *energy demand* | – |
| *air qual* | – | *biodivers* | – | *energy recoveri* | – |
| *urban agglomer* | 0.935 | *urban sustain* | 0.946 | *wte* | 0.953 |
| *ssp* | 0.923 | *green growth* | 0.935 | *waste inciner* | 0.946 |
| *scale effect* | 0.920 | *synergi* | 0.934 | *rdf* | 0.924 |
| *power sector* | 0.919 | *social sustain* | 0.932 | *inciner* | 0.920 |
| *co benefit* | 0.919 | *resili* | 0.932 | *material recoveri* | 0.903 |
| *drought* | 0.918 | *adaptive capac* | 0.931 | *landfill ga* | 0.901 |
| *ekc* | 0.917 | *water scarc* | 0.930 | *environmental benefit* | 0.894 |
| *economic structur* | 0.914 | *ecological network analysi* | 0.929 | *waste treat* | 0.886 |
| *coal consumpt* | 0.913 | *food secur* | 0.926 | *mrf* | 0.863 |
| *carbon intens* | 0.911 | *co benefit* | 0.926 | *mswi* | 0.861 |
| *emission reduct* | 0.910 | *water manag* | 0.925 | *energy intens* | 0.859 |
| *economic develop* | 0.909 | *food energy water nexus* | 0.917 | *gasif* | 0.859 |
| *mrio analysi* | 0.908 | *water secur* | 0.916 | *organic wast* | 0.858 |
| *regional differ* | 0.908 | *sdg* | 0.916 | *open dump* | 0.857 |
| *dewat* | 0.904 | *equiti* | 0.916 | *cement kiln* | 0.857 |
| *hdi* | 0.903 | *urban wat* | 0.915 | *primary energi* | 0.854 |
| *driving forc* | 0.903 | *nexus* | 0.915 | *electricity gener* | 0.854 |
| *policy impl* | 0.902 | *policy impl* | 0.912 | *pollution emiss* | 0.851 |
| *energy secur* | 0.898 | *urban agglomer* | 0.912 | *biowast* | 0.849 |
| *industrial transform* | 0.897 | *water energy nexus* | 0.910 | *energy sav* | 0.846 |
| *biodivers* | 0.897 | *eco innov* | 0.909 | *chemical recycl* | 0.844 |
| *decoupl* | 0.896 | *tourism* | 0.908 | *mechanical recycl* | 0.843 |
| *ammonia* | 0.896 | *green develop* | 0.908 | *co process* | 0.842 |
| *carbon neutr* | 0.896 | *urban metabol* | 0.908 | *biodegradable wast* | 0.841 |
| *tourism* | 0.894 | *hdi* | 0.907 | *techno economic assess* | 0.837 |
| *decontamin* | 0.892 | *lean* | 0.906 | *anaerobic digest* | 0.834 |
| *socioeconomic factor* | 0.892 | *green supply chain* | 0.906 | *bioethanol* | 0.834 |
| *ion exchang* | 0.890 | *sustainable consumption and product* | 0.905 | *municipal wast* | 0.831 |
| *ecological secur* | 0.890 | *chemical industrial park* | 0.905 | *economic cost* | 0.831 |
| *economic growth* | 0.890 | *emerging economi* | 0.904 | *ghg* | 0.826 |
| *emission inventori* | 0.889 | *energy secur* | 0.902 | *material product* | 0.825 |
| *lmdi* | 0.889 | *green build* | 0.900 | *shale ga* | 0.823 |
| *water energy nexus* | 0.889 | *energy manag* | 0.900 | *pla* | 0.821 |
| *carrying capac* | 0.888 | *network analysi* | 0.897 | *mbt* | 0.821 |
| *phosphorus remov* | 0.888 | *air qual* | 0.897 | *embodied energi* | 0.821 |
| *water scarc* | 0.887 | *technological innov* | 0.895 | *neat* | 0.816 |
| *green develop* | 0.887 | *sharing economi* | 0.895 | *sanitary landfil* | 0.815 |
| *energy structur* | 0.886 | *human develop* | 0.894 | *land use chang* | 0.813 |
| *carbon dioxide intens* | 0.886 | *supplier select* | 0.894 | *organic solid wast* | 0.813 |
| *water treat* | 0.885 | *conceptual framework* | 0.894 | *green wast* | 0.813 |
| *solubl* | 0.884 | *nutrient flow* | 0.892 | *power gener* | 0.813 |
| *rural* | 0.884 | *sustainable water manag* | 0.891 | *efw* | 0.812 |
| *pollution emiss* | 0.884 | *urban plan* | 0.891 | *cradle to g* | 0.811 |
| *demand respons* | 0.882 | *environmental qu* | 0.890 | *economic assess* | 0.811 |
| *leachat* | 0.882 | *urban wat* | 0.889 | *hydropow* | 0.810 |
| *green growth* | 0.882 | *sustainable solid waste manag* | 0.888 | *energy effici* | 0.807 |
| *activated sludg* | 0.880 | *urban water manag* | 0.887 | *land appl* | 0.805 |
| *water secur* | 0.879 | *electronic commerc* | 0.887 | *district h* | 0.805 |
| *urban wat* | 0.879 | *singapor* | 0.886 | *carbon footprint* | 0.805 |
| *water retent* | 0.879 | *operational excel* | 0.886 | *carbon emiss* | 0.805 |
| *residential electricity consumpt* | 0.878 | *urban infrastructur* | 0.886 | *global warm* | 0.804 |
| *influencing factor* | 0.878 | *energy conserv* | 0.885 | *biological treat* | 0.804 |
| *immobil* | 0.878 | *climate change mitig* | 0.885 | *climate impact* | 0.803 |
| *tfp* | 0.878 | *tbl* | 0.885 | *carbon bal* | 0.803 |
| *decomposition analysi* | 0.877 | *carrying capac* | 0.884 | *thermal treat* | 0.803 |
| *carbon reduct* | 0.876 | *scale effect* | 0.884 | *fossil fuel* | 0.800 |
| *socioeconomic driv* | 0.875 | *social metabol* | 0.883 | *landfil* | 0.800 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Food | | Land | | Metals | |
| term | sim | term | sim | term | sim |
| *food* | – | *land* | – | *metal* | – |
| *food wast* | – | *land us* | – | *metal recycl* | – |
| *food loss* | – | *land footprint* | – | *hydrometallurgi* | – |
| *food supply chain* | 0.948 | *carbon sequestr* | 0.965 | *spent catalyst* | 0.977 |
| *restaur* | 0.921 | *particulate matt* | 0.951 | *hydrometallurgical process* | 0.972 |
| *supermarket* | 0.874 | *greenhouse gas inventori* | 0.951 | *pyrometallurgi* | 0.967 |
| *food servic* | 0.866 | *hydropow* | 0.950 | *pyrometallurgical process* | 0.965 |
| *food waste reduct* | 0.863 | *mitigation potenti* | 0.948 | *silicon* | 0.961 |
| *food secur* | 0.856 | *socioeconomic driv* | 0.948 | *lead acid batteri* | 0.961 |
| *food industri* | 0.846 | *process emiss* | 0.947 | *wpcb* | 0.958 |
| *lifestyl* | 0.843 | *environmental account* | 0.943 | *organic acid* | 0.957 |
| *healthcar* | 0.840 | *soybean oil* | 0.943 | *carbon fib* | 0.957 |
| *united kingdom* | 0.840 | *premature mort* | 0.943 | *valuable met* | 0.956 |
| *food system* | 0.837 | *embodied emiss* | 0.942 | *shredder residu* | 0.950 |
| *consumer behavior* | 0.835 | *agricultural sustain* | 0.942 | *bioleach* | 0.947 |
| *food product* | 0.834 | *water use effici* | 0.941 | *precious met* | 0.947 |
| *plastic pollut* | 0.831 | *mrio analysi* | 0.940 | *aluminum dross* | 0.945 |
| *tourism* | 0.829 | *hydropon* | 0.937 | *bauxite residu* | 0.940 |
| *food recoveri* | 0.825 | *cmaq* | 0.937 | *lcd* | 0.933 |
| *electronic commerc* | 0.823 | *macrophyt* | 0.935 | *solvent extract* | 0.932 |
| *canteen* | 0.820 | *carbon emission intens* | 0.935 | *asr* | 0.931 |
| *respons* | 0.820 | *ecological secur* | 0.935 | *shredder* | 0.928 |
| *climate impact* | 0.819 | *ceen* | 0.933 | *flame retard* | 0.927 |
| *developing economi* | 0.819 | *molass* | 0.932 | *recovery of met* | 0.923 |
| *veget* | 0.817 | *nitrogen use effici* | 0.931 | *coal ash* | 0.922 |
| *thailand* | 0.817 | *nutrient bal* | 0.930 | *galvan* | 0.921 |
| *canada* | 0.810 | *coal fired power pl* | 0.928 | *lithium recycl* | 0.920 |
| *waste prevent* | 0.810 | *urban heat island* | 0.927 | *crt* | 0.920 |
| *furnitur* | 0.807 | *chemical process* | 0.927 | *catalyst* | 0.920 |
| *sustainable consumption and product* | 0.805 | *carbon bal* | 0.923 | *pgm* | 0.919 |
| *bioeconomi* | 0.803 | *gas emiss* | 0.922 | *pcb* | 0.918 |
| *mass flow analysi* | 0.802 | *potable water sav* | 0.922 | *catalytic convert* | 0.915 |
| *netherland* | 0.800 | *technology adopt* | 0.922 | *polyvinyl chlorid* | 0.912 |
| *economic impact* | 0.800 | *lmdi* | 0.921 | *thermoplast* | 0.911 |
| *bike shar* | 0.800 | *evapotranspir* | 0.921 | *lead slag* | 0.910 |
| *south africa* | 0.799 | *combined heat and pow* | 0.921 | *basel convent* | 0.910 |
| *retail* | 0.796 | *energy convers* | 0.920 | *critical met* | 0.909 |
| *healthcare wast* | 0.795 | *peak miner* | 0.920 | *recycling process* | 0.908 |
| *resource scarc* | 0.795 | *low impact develop* | 0.920 | *waste plast* | 0.908 |
| *procur* | 0.794 | *water balance model* | 0.920 | *polyolefin* | 0.908 |
| *farm* | 0.792 | *wastewater minim* | 0.920 | *reprocess* | 0.906 |
| *sdg* | 0.792 | *land use chang* | 0.919 | *vacuum* | 0.906 |
| *food packag* | 0.792 | *consequential model* | 0.918 | *cathode materi* | 0.905 |
| *latin america* | 0.792 | *water secur* | 0.918 | *crm* | 0.904 |
| *biodivers* | 0.790 | *herbaceous mimosa* | 0.918 | *supercritical co2* | 0.901 |
| *malaysia* | 0.790 | *alfalfa* | 0.917 | *preprocess* | 0.900 |
| *singapor* | 0.788 | *ethanol product* | 0.917 | *wastewater sludge ash* | 0.899 |
| *climate change mitig* | 0.786 | *slaughterhous* | 0.916 | *stainless steel* | 0.899 |
| *medical wast* | 0.785 | *consumption based account* | 0.916 | *wood wast* | 0.899 |
| *energy manag* | 0.785 | *cge* | 0.915 | *polystyren* | 0.896 |
| *consumption pattern* | 0.784 | *forestri* | 0.915 | *polypropylen* | 0.895 |
| *bmw* | 0.784 | *organic municipal solid wast* | 0.914 | *biotechnolog* | 0.894 |
| *indonesia* | 0.784 | *marin* | 0.913 | *aluminum alloy* | 0.894 |
| *portug* | 0.783 | *spa* | 0.913 | *polyest* | 0.893 |
| *prevent* | 0.783 | *nitrogen remov* | 0.912 | *mswi bottom ash* | 0.893 |
| *european union* | 0.782 | *ekc* | 0.912 | *gallium* | 0.893 |
| *sri lanka* | 0.781 | *stochastic frontier analysi* | 0.912 | *waste util* | 0.892 |
| *paper recycl* | 0.780 | *sustainable water us* | 0.912 | *mechanical process* | 0.892 |
| *fruit* | 0.780 | *regional differ* | 0.911 | *reclam* | 0.892 |
| *global supply chain* | 0.780 | *ssp* | 0.911 | *industrial wast* | 0.891 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Minerals | | Waste | | Water | | Wood | |
| term | sim | term | sim | term | sim | term | sim |
| *miner* | – | *wast* | – | *water* | – | *wood* | – |
| *mineral resourc* | – | *waste manag* | – | *wastewat* | – | *wood product* | – |
| *mineral deplet* | – | *solid wast* | – | *rainwater harvest* | – | *recovered wood* | – |
| *non fuel miner* | 0.992 | *integrated waste manag* | 0.884 | *rainwater harvesting system* | 0.970 | *energy cont* | 0.971 |
| *population balance model* | 0.985 | *swm* | 0.870 | *water suppli* | 0.929 | *biomass ash* | 0.970 |
| *energy scenario* | 0.981 | *waste recoveri* | 0.868 | *rainwater tank* | 0.917 | *building stock* | 0.966 |
| *peak miner* | 0.981 | *mswm* | 0.865 | *water demand* | 0.893 | *nanomateri* | 0.965 |
| *time series analysi* | 0.981 | *waste reduct* | 0.859 | *wastewater treat* | 0.885 | *ferronickel* | 0.964 |
| *consequential lca* | 0.980 | *sanitary landfil* | 0.850 | *rainwat* | 0.885 | *waste util* | 0.963 |
| *sustainable build* | 0.979 | *waste polici* | 0.850 | *water sav* | 0.878 | *thermodynamic rar* | 0.962 |
| *anthropogenic resourc* | 0.979 | *municipal wast* | 0.849 | *effluent* | 0.871 | *timber* | 0.961 |
| *copper consumpt* | 0.979 | *dar es salaam* | 0.846 | *irrig* | 0.855 | *recycled pap* | 0.961 |
| *stock dynam* | 0.978 | *selective collect* | 0.846 | *water reus* | 0.854 | *ethanol product* | 0.961 |
| *computer recycl* | 0.977 | *epr* | 0.840 | *potable water sav* | 0.846 | *aluminum alloy* | 0.960 |
| *delphi studi* | 0.977 | *zero wast* | 0.836 | *urban wat* | 0.846 | *plasterboard* | 0.958 |
| *policy simul* | 0.977 | *open dump* | 0.834 | *reliabl* | 0.842 | *exergy replacement cost* | 0.958 |
| *mrio analysi* | 0.976 | *urban wast* | 0.832 | *tank siz* | 0.839 | *papermak* | 0.957 |
| *exergy replacement cost* | 0.976 | *informal sector* | 0.830 | *wwtp* | 0.833 | *post consumer wast* | 0.956 |
| *ccu* | 0.975 | *wale* | 0.828 | *water bal* | 0.828 | *scrap qual* | 0.954 |
| *complex system* | 0.975 | *resource conserv* | 0.828 | *water manag* | 0.825 | *ion adsorption clay* | 0.953 |
| *copper recycl* | 0.975 | *belgium* | 0.827 | *climatic condit* | 0.820 | *metal finish* | 0.953 |
| *metal cycl* | 0.975 | *tanzania* | 0.825 | *groundwat* | 0.820 | *calcium carbon* | 0.951 |
| *spa* | 0.974 | *waste character* | 0.821 | *water resourc* | 0.816 | *metal cycl* | 0.951 |
| *consumption based account* | 0.974 | *hazardous wast* | 0.818 | *wastewater reus* | 0.811 | *wood wast* | 0.951 |
| *decarbonis* | 0.973 | *municipal waste manag* | 0.817 | *graywater reus* | 0.809 | *rerefin* | 0.950 |
| *regional approach* | 0.973 | *biodegradable wast* | 0.816 | *urban wat* | 0.809 | *shredder residu* | 0.949 |
| *lcsa* | 0.973 | *msw* | 0.815 | *water conserv* | 0.809 | *ceen* | 0.949 |
| *rerefin* | 0.972 | *switzerland* | 0.811 | *graywat* | 0.807 | *mechanical process* | 0.949 |
| *data driven approach* | 0.972 | *packaging wast* | 0.811 | *water scarc* | 0.807 | *pyrometallurgi* | 0.949 |
| *mass flow analysi* | 0.971 | *sustainable waste manag* | 0.810 | *treatment* | 0.806 | *gas emiss* | 0.949 |
| *colombia* | 0.971 | *developing countri* | 0.810 | *airport* | 0.802 | *engineering appl* | 0.948 |
| *interval program* | 0.971 | *recycling system* | 0.810 | *blackwat* | 0.801 | *thermochemical convers* | 0.948 |
| *numerical model* | 0.970 | *electronic waste manag* | 0.809 | *water qu* | 0.798 | *spent automotive catalyst* | 0.947 |
| *structural analysi* | 0.970 | *ireland* | 0.805 | *drought* | 0.790 | *spent coffee ground* | 0.947 |
| *product develop* | 0.970 | *indonesia* | 0.805 | *mediterranean* | 0.787 | *recovery of met* | 0.946 |
| *hard disk dr* | 0.969 | *household wast* | 0.803 | *green infrastructur* | 0.779 | *alkali activated materi* | 0.946 |
| *demateri* | 0.969 | *sri lanka* | 0.802 | *rainfall time seri* | 0.778 | *used oil* | 0.945 |
| *chemical process* | 0.969 | *solid waste recycl* | 0.800 | *water secur* | 0.777 | *foundry sand* | 0.945 |
| *bayesian network* | 0.968 | *waste collect* | 0.800 | *water network* | 0.776 | *consequential lca* | 0.944 |
| *material crit* | 0.968 | *landfill direct* | 0.799 | *sustainable water manag* | 0.774 | *waste wood* | 0.944 |
| *business perform* | 0.968 | *environmental manag* | 0.798 | *struvit* | 0.772 | *resource sav* | 0.943 |
| *consequential model* | 0.968 | *greec* | 0.797 | *dea* | 0.770 | *soil amend* | 0.943 |
| *scrap qual* | 0.967 | *wte* | 0.797 | *synergi* | 0.770 | *alternative materi* | 0.943 |
| *process emiss* | 0.967 | *integrated solid waste manag* | 0.795 | *harvest* | 0.768 | *tensile properti* | 0.943 |
| *nitrogen use effici* | 0.967 | *european union* | 0.794 | *stochast* | 0.766 | *polymer recycl* | 0.941 |
| *shared mobl* | 0.967 | *critical success factor* | 0.794 | *recycled wat* | 0.765 | *utilization r* | 0.941 |
| *image recognit* | 0.967 | *waste prevent* | 0.793 |  |  | *dynamic mfa* | 0.939 |
| *combined heat and pow* | 0.966 | *municip* | 0.793 |  |  | *soybean oil* | 0.938 |
| *environmental account* | 0.966 | *urban solid wast* | 0.793 |  |  | *incineration ash* | 0.938 |
| *social life cycle assess* | 0.965 | *portug* | 0.790 |  |  | *gas chromatographi* | 0.937 |
| *patent* | 0.964 | *dispos* | 0.790 |  |  | *shredder* | 0.937 |
| *chance constrained program* | 0.963 | *waste dispos* | 0.790 |  |  | *paper sludg* | 0.937 |
| *waste trad* | 0.963 | *poland* | 0.789 |  |  | *soil condition* | 0.937 |
| *decoupling analysi* | 0.963 | *economic instru* | 0.789 |  |  | *semiconductor* | 0.936 |
| *analytic network process* | 0.962 | *bibliometric analysi* | 0.789 |  |  | *anthropogenic resourc* | 0.936 |
| *stochastic frontier analysi* | 0.962 | *resource recoveri* | 0.788 |  |  | *supercritical co2* | 0.935 |
| *agricultural sustain* | 0.962 | *malaysia* | 0.786 |  |  | *environmental tax* | 0.935 |
| *environmental tax* | 0.962 | *czech republ* | 0.786 |  |  | *spent lithium ion batteri* | 0.935 |
| *british columbia* | 0.961 | *3r* | 0.785 |  |  | *gfrp* | 0.935 |
| *stakeholder analysi* | 0.961 | *taiwan* | 0.785 |  |  | *beneficial us* | 0.935 |

Table S12. Summary of resource-surrogates (terminology and similarity) that were identified based on the word2vec-RCR model and expert knowledge for the 10 resources (ordered by the number of surrogates from low to high).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Air | | Ecosystem | | Land | | Water | |
| term | sim | term | sim | term | sim | term | sim |
| *air qual* | 0.917 | *few nexus* | 0.957 | *carbon sequestr* | 0.965 | *rainwater harvesting system* | 0.970 |
| *air pollut* | 0.917 | *green growth* | 0.935 | *forest land* | 0.959 | *water suppli* | 0.929 |
| *emission reduct* | 0.910 | *resili* | 0.932 | *land area* | 0.957 | *rainwater tank* | 0.917 |
| *emission inventori* | 0.889 | *adaptive capac* | 0.931 | *agricultural land occup* | 0.950 | *water demand* | 0.893 |
| *carbon dioxide intens* | 0.886 | *ecological network analysi* | 0.929 | *arable land* | 0.948 | *wastewater treat* | 0.885 |
| *pollution emiss* | 0.884 | *ecosystem servic* | 0.927 | *soybean oil* | 0.943 | *rainwat* | 0.885 |
| *nitrous oxid* | 0.870 | *biodivers* | 0.927 | *marginal land* | 0.942 | *water sav* | 0.878 |
| *carbon emission intens* | 0.864 | *food secur* | 0.926 | *agricultural land* | 0.942 | *effluent* | 0.871 |
| *severe air pollut* | 0.855 | *food energy water nexus* | 0.917 | *agricultural sustain* | 0.942 | *irrig* | 0.855 |
| *ozon* | 0.851 | *nexus* | 0.915 | *forestland* | 0.939 | *water reus* | 0.854 |
| *air pollution emiss* | 0.839 | *water energy nexus* | 0.910 | *peatland* | 0.938 | *potable water sav* | 0.846 |
| *carbon dioxide sequestr* | 0.838 | *eco innov* | 0.909 | *land conserv* | 0.935 | *urban wat* | 0.846 |
| *airport co2* | 0.834 | *green develop* | 0.908 | *land occup* | 0.934 | *tank siz* | 0.839 |
| *methan* | 0.834 | *urban metabol* | 0.908 | *unused land* | 0.933 | *wwtp* | 0.833 |
| *pm2.5* | 0.833 | *ecosystem* | 0.904 | *land demand* | 0.931 | *water bal* | 0.828 |
| *airborne diseas* | 0.827 | *carrying capac* | 0.884 | *cultivated land* | 0.929 | *watersh* | 0.825 |
| *cmaq* | 0.819 | *food system* | 0.881 | *land cov* | 0.926 | *water manag* | 0.825 |
| *carbon emiss* | 0.819 | *ecological secur* | 0.870 | *farmland* | 0.923 | *groundwat* | 0.820 |
| *sulfur dioxid* | 0.818 | *sustainable agricultur* | 0.867 | *land use chang* | 0.919 | *water resourc* | 0.816 |
| *odor* | 0.817 | *urban ecolog* | 0.866 | *land dispos* | 0.918 | *wastewater reus* | 0.811 |
| *hydrogen* | 0.814 | *eco effici* | 0.860 | *extended land consumpt* | 0.914 | *graywater reus* | 0.809 |
| *airborn* | 0.811 | *ekc* | 0.858 | *maiz* | 0.911 | *urban water system* | 0.809 |
| *ccus* | 0.807 | *urban ecosystem* | 0.852 | *sustainable agricultur* | 0.910 | *water conserv* | 0.809 |
| *embodied emiss* | 0.798 | *ecosystem restor* | 0.851 | *rice* | 0.904 | *graywat* | 0.807 |
| *emiss* | 0.797 | *forest ecosystem* | 0.834 | *crop residu* | 0.904 | *water scarc* | 0.807 |
| *process emiss* | 0.776 | *ecosystem health* | 0.829 | *water and land us* | 0.901 | *blackwat* | 0.801 |
| *ccu* | 0.774 | *natural ecosystem* | 0.828 | *landfarm* | 0.898 | *water qu* | 0.798 |
| *air pollution control* | 0.768 | *ecosystem service valu* | 0.828 | *soil fertil* | 0.896 | *drought* | 0.790 |
|  |  | *agroecosystem* | 0.822 | *land resourc* | 0.894 | *water stress* | 0.789 |
|  |  | *green infrastructur* | 0.815 | *grassland* | 0.893 | *water system* | 0.788 |
|  |  | *ecosystem qu* | 0.812 | *farmland protect* | 0.891 | *water alloc* | 0.788 |
|  |  | *ecosystem services valu* | 0.811 | *urban agglomer* | 0.877 | *water util* | 0.783 |
|  |  | *resource manag* | 0.800 | *cropland* | 0.827 | *rainfall time seri* | 0.778 |
|  |  | *natural resourc* | 0.784 | *land footprint* | 0.789 | *water secur* | 0.777 |
|  |  | *resource product* | 0.773 | *crop* | 0.787 | *water network* | 0.776 |
|  |  |  |  | *agricultur* | 0.786 | *sustainable water manag* | 0.774 |
|  |  |  |  | *farm* | 0.775 | *struvit* | 0.772 |
|  |  |  |  |  |  | *water sourc* | 0.771 |
|  |  |  |  |  |  | *water pric* | 0.767 |
|  |  |  |  |  |  | *recycled wat* | 0.765 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Minerals | | Wood | | Energy | | Food | |
| term | sim | term | sim | term | sim | term | sim |
| *non fuel miner* | 0.992 | *industrial wood* | 0.968 | *wte* | 0.953 | *food supply chain* | 0.948 |
| *peak miner* | 0.981 | *wood pap* | 0.966 | *rdf* | 0.924 | *restaur* | 0.921 |
| *anthropogenic resourc* | 0.979 | *wood fib* | 0.965 | *inciner* | 0.920 | *food consumpt* | 0.878 |
| *mineral commod* | 0.979 | *urban wood wast* | 0.965 | *energy intens* | 0.859 | *supermarket* | 0.874 |
| *construction miner* | 0.978 | *woody biomass* | 0.964 | *gasif* | 0.859 | *food servic* | 0.866 |
| *computer recycl* | 0.977 | *wooden* | 0.963 | *primary energi* | 0.854 | *food loss* | 0.866 |
| *biominer* | 0.971 | *woody biomass ash* | 0.963 | *electricity gener* | 0.854 | *food* | 0.866 |
| *mass flow analysi* | 0.971 | *timber* | 0.961 | *energy sav* | 0.846 | *food waste reduct* | 0.863 |
| *non metallic miner* | 0.971 | *recycled pap* | 0.961 | *incineration with energy recoveri* | 0.845 | *food waste gener* | 0.858 |
| *hard disk dr* | 0.969 | *firewood* | 0.959 | *bioethanol* | 0.834 | *food secur* | 0.856 |
| *demateri* | 0.969 | *plasterboard* | 0.958 | *energy consum* | 0.832 | *food manufactur* | 0.853 |
| *material crit* | 0.968 | *papermak* | 0.957 | *shale ga* | 0.823 | *food industri* | 0.846 |
| *ewmfa* | 0.961 | *wood pellet* | 0.955 | *embodied energi* | 0.821 | *food system* | 0.837 |
| *mineral wast* | 0.961 | *wood bas* | 0.953 | *power gener* | 0.813 | *food product* | 0.834 |
| *mineral deposit* | 0.960 | *wood cascad* | 0.952 | *hydropow* | 0.810 | *food wast* | 0.833 |
| *scarciti* | 0.952 | *wood wast* | 0.951 | *energy recovery effici* | 0.808 | *staple food* | 0.829 |
| *supply risk* | 0.943 | *wood based product* | 0.947 | *energy effici* | 0.807 | *global food* | 0.828 |
| *critical materi* | 0.942 | *wood polym* | 0.947 | *district h* | 0.805 | *food recoveri* | 0.825 |
| *non renewable resourc* | 0.928 | *wood based materi* | 0.947 | *fossil fuel* | 0.800 | *agri food* | 0.822 |
| *resource deplet* | 0.922 | *woodfibr* | 0.947 | *energy requir* | 0.799 | *canteen* | 0.820 |
| *mine* | 0.888 | *harvested wood product* | 0.944 | *primary energy us* | 0.799 | *veget* | 0.817 |
| *critic* | 0.888 | *waste wood* | 0.944 | *energy product* | 0.797 | *food safeti* | 0.816 |
| *dissip* | 0.874 | *wood chip* | 0.941 | *energy sourc* | 0.796 | *food suppli* | 0.812 |
| *ore grad* | 0.864 | *fuelwood* | 0.938 | *renewable energi* | 0.794 | *food deliveri* | 0.811 |
| *dynamic mfa* | 0.860 | *paper sludg* | 0.937 | *wind pow* | 0.794 | *wasted food* | 0.806 |
| *urban min* | 0.856 | *wood us* | 0.936 | *power sector* | 0.794 | *retail food* | 0.803 |
| *resource scarc* | 0.852 | *harvested wood* | 0.931 | *natural ga* | 0.792 | *food waste manag* | 0.802 |
| *phosphorus rock* | 0.846 | *wood residu* | 0.927 | *energi* | 0.792 | *household food* | 0.802 |
| *landfill min* | 0.844 | *woodi* | 0.927 | *energy demand* | 0.792 | *edible food* | 0.797 |
| *mineral deplet* | 0.834 | *wood cut* | 0.924 | *energy structur* | 0.789 | *food waste quantif* | 0.797 |
| *material flow* | 0.833 | *wood waste ash* | 0.913 | *total energi* | 0.788 | *food rel* | 0.795 |
| *material consumpt* | 0.817 | *wood fly ash* | 0.912 | *energy resourc* | 0.787 | *food surplus* | 0.793 |
| *mineral process* | 0.812 | *wood product* | 0.910 | *national energi* | 0.787 | *food item* | 0.793 |
| *mineral construct* | 0.812 | *use of wood* | 0.902 | *fossil energi* | 0.785 | *global food wast* | 0.793 |
| *miner* | 0.795 | *wood resourc* | 0.897 | *energy system* | 0.785 | *farm* | 0.792 |
| *mineral resourc* | 0.795 | *paper recycl* | 0.877 | *energy cont* | 0.784 | *food packag* | 0.792 |
| *phosphat* | 0.787 | *paper industri* | 0.877 | *energy mix* | 0.783 | *food related routin* | 0.792 |
| *potassium* | 0.780 | *waste wood manag* | 0.876 | *electr* | 0.779 | *food waste recycl* | 0.787 |
| *weibull distribut* | 0.779 | *wood* | 0.872 | *exergy analysi* | 0.779 | *food group* | 0.787 |
| *silicon* | 0.779 | *recovered wood* | 0.872 | *energy sector* | 0.778 | *food loss and wast* | 0.785 |
| *phosphor* | 0.772 | *pulp and paper industri* | 0.871 | *energy suppli* | 0.778 | *fruit* | 0.780 |
|  |  | *wood fram* | 0.836 | *energy rel* | 0.774 | *plate wast* | 0.776 |
|  |  |  |  | *energy gener* | 0.770 | *food demand* | 0.774 |
|  |  |  |  | *solar energi* | 0.767 | *food waste problem* | 0.771 |
|  |  |  |  |  |  | *nutrit* | 0.771 |
|  |  |  |  |  |  | *refriger* | 0.771 |
|  |  |  |  |  |  | *food value chain* | 0.770 |
|  |  |  |  |  |  | *food waste diari* | 0.768 |
|  |  |  |  |  |  | *avoidable food wast* | 0.767 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Waste | | Metals | | | |
| term | sim | term | sim | term | sim |
| *integrated waste manag* | 0.884 | *hydrometallurg* | 0.974 | *metal ion* | 0.852 |
| *swm* | 0.870 | *pyrometallurg* | 0.973 | *gold* | 0.852 |
| *waste management strategi* | 0.869 | *hydrometallurgical process* | 0.972 | *metal product* | 0.850 |
| *waste recoveri* | 0.868 | *metal recycl* | 0.967 | *platinum* | 0.847 |
| *mswm* | 0.865 | *hydrometallurgi* | 0.967 | *magnesium* | 0.841 |
| *waste reduct* | 0.859 | *pyrometallurgi* | 0.967 | *metalloid* | 0.835 |
| *sanitary landfil* | 0.850 | *pyrometallurgical process* | 0.965 | *tin* | 0.832 |
| *waste polici* | 0.850 | *lead acid batteri* | 0.961 | *metal cycl* | 0.831 |
| *municipal wast* | 0.849 | *metallurg* | 0.958 | *corros* | 0.827 |
| *zero wast* | 0.836 | *wpcb* | 0.958 | *biohydrometallurgi* | 0.822 |
| *open dump* | 0.834 | *valuable met* | 0.956 | *iron* | 0.821 |
| *urban wast* | 0.832 | *bioleach* | 0.947 | *metallic fract* | 0.817 |
| *informal sector* | 0.830 | *precious met* | 0.947 | *chromium* | 0.817 |
| *waste character* | 0.821 | *aluminum dross* | 0.945 | *scarce met* | 0.813 |
| *hazardous wast* | 0.818 | *bauxite residu* | 0.940 | *metallurgical recoveri* | 0.811 |
| *municipal waste manag* | 0.817 | *metal valu* | 0.939 | *metal tox* | 0.811 |
| *biodegradable wast* | 0.816 | *metal cont* | 0.924 | *mercuri* | 0.810 |
| *msw* | 0.815 | *recovery of met* | 0.923 | *metal* | 0.808 |
| *waste hierarchi* | 0.813 | *galvan* | 0.921 | *aluminum recycl* | 0.808 |
| *packaging wast* | 0.811 | *lithium recycl* | 0.920 | *potassium* | 0.807 |
| *sustainable waste manag* | 0.810 | *pcb* | 0.918 | *pyrometallurgy and hydrometallurgi* | 0.806 |
| *waste management plan* | 0.810 | *lead slag* | 0.910 | *metal electrolysi* | 0.806 |
| *recycling system* | 0.810 | *critical met* | 0.909 | *spent lead acid batteri* | 0.803 |
| *electronic waste manag* | 0.809 | *metal resourc* | 0.902 | *lightweight metal scrap* | 0.803 |
| *household wast* | 0.803 | *stainless steel* | 0.899 | *metal contamin* | 0.801 |
| *solid waste recycl* | 0.800 | *metallurgi* | 0.895 | *metal industri* | 0.791 |
| *waste collect* | 0.800 | *aluminum alloy* | 0.894 | *metal packag* | 0.791 |
| *landfill direct* | 0.799 | *gallium* | 0.893 | *copper consumpt* | 0.789 |
| *wte* | 0.797 | *pyro hydrometallurg* | 0.889 | *metal demand* | 0.788 |
| *integrated solid waste manag* | 0.795 | *spent lithium ion batteri* | 0.888 | *metallifer* | 0.786 |
| *waste prevent* | 0.793 | *ferronickel* | 0.886 | *aluminum* | 0.786 |
| *municip* | 0.793 | *metal extract* | 0.886 | *lithium* | 0.785 |
| *urban solid wast* | 0.793 | *metal finish* | 0.878 | *silver* | 0.783 |
| *dispos* | 0.790 | *ferrous scrap* | 0.877 | *trace met* | 0.783 |
| *waste dispos* | 0.790 | *metallurgical industri* | 0.872 | *tan* | 0.782 |
| *waste management system* | 0.789 | *hydrometallurgical recoveri* | 0.871 | *copper recycl* | 0.780 |
| *3r* | 0.785 | *lithium ion batteri* | 0.871 | *nickel* | 0.777 |
| *plastic packaging wast* | 0.782 | *metallic el* | 0.869 | *metal leach* | 0.775 |
| *healthcare wast* | 0.781 | *metallurgical process* | 0.868 | *zinc* | 0.774 |
| *waste divers* | 0.777 | *copper slag* | 0.865 | *tellurium* | 0.767 |
| *recyclable materi* | 0.777 | *alloying el* | 0.862 |  |  |
| *waste composit* | 0.776 | *non ferrous metallurgi* | 0.860 |  |  |
| *organic solid wast* | 0.775 | *nickel metal hydrid* | 0.859 |  |  |
| *municipal solid waste collect* | 0.774 | *hydrometallurgical leach* | 0.858 |  |  |
| *waste management practic* | 0.774 | *metallurgical recycl* | 0.857 |  |  |
| *organic wast* | 0.773 | *metal scrap* | 0.857 |  |  |
| *waste sector* | 0.767 | *household metal packag* | 0.855 |  |  |
| *informal recycl* | 0.765 | *hydro metallurg* | 0.855 |  |  |
| *gypsum wast* | 0.765 | *metal bear* | 0.854 |  |  |

**Text S6. Optimal *q* and threshold values for classification accuracy**

An assessment was subsequently taken to evaluate overall classification accuracy based on different thresholds and *q* values (Fig. S9). Higher accuracies are found at different combinations of *q* values and threshold values, and the highest weighted accuracy (≈ 0.74) is identified based on a threshold of 0.85 and a *q* value of 178. In other words, for each paper and each resource that the top 178 highest similarity values were used to evaluate its publication similarity to a resource, which was used to determine if the paper closely focused on the resource (publication similarity ≥ 0.85) or not.

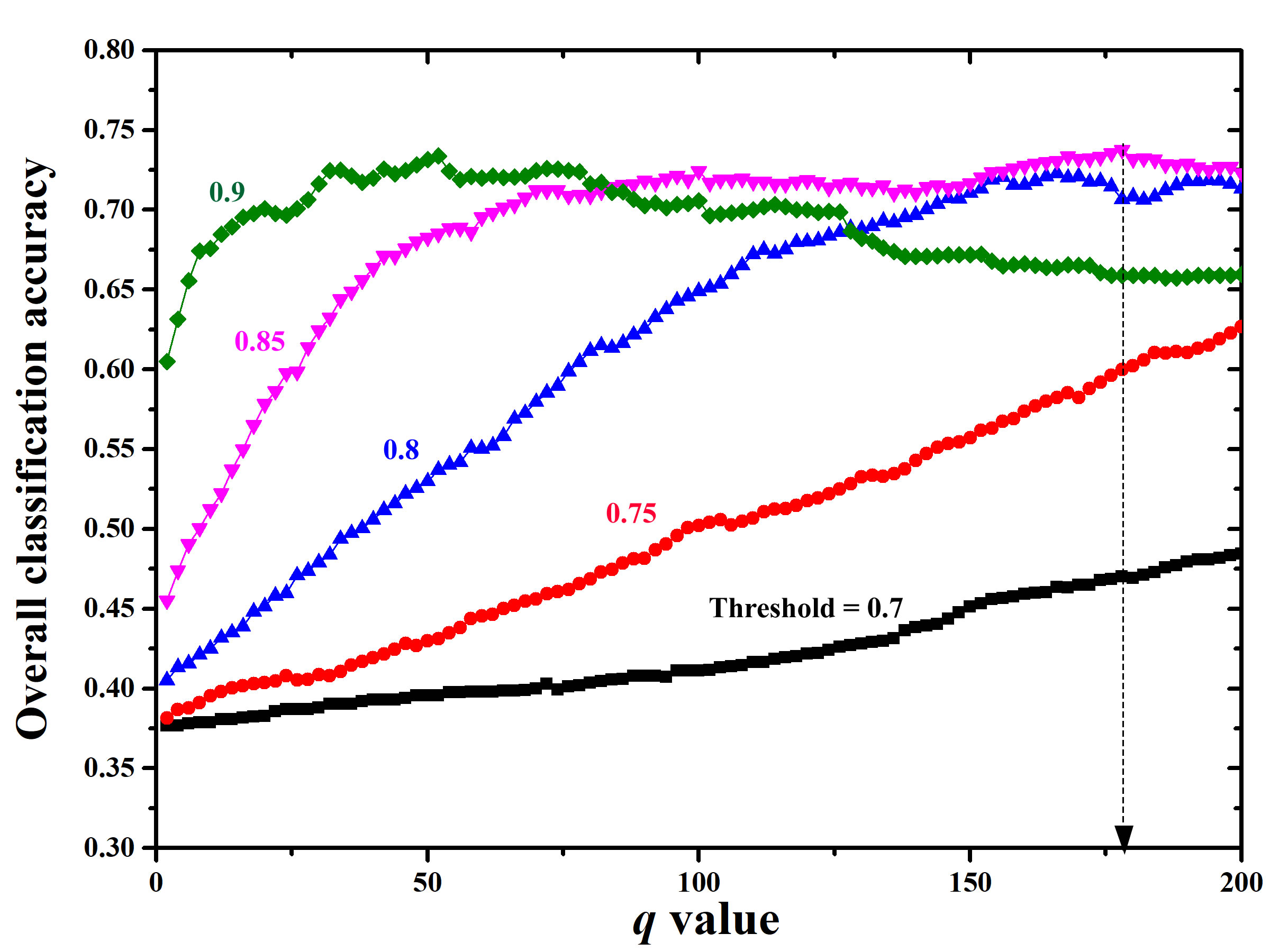


Fig. S9. Overall classification accuracy of the 112 reviewed papers based on different threshold and q values. The highest weighted accuracy was identified based on a threshold of 0.85 and a *q* value of 178.

**References**

Gensim (2022). Gensim is a library for various natural language processing functionalities based on modern statistical machine learning. <https://radimrehurek.com/gensim/>

Hinton, G. E., & Roweis, S. (2002). Stochastic neighbor embedding. Advances in neural information processing systems, 15.

Sklearn. (2022). Scikit-learn is a free software machine learning library for the Python programming language. <https://scikit-learn.org/>.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).